collected and recorded from the time when the patient signed the informed consent until the study completion. Overall, during the initial treatment phase of the study, 56.9 percent patients of Synvisc-One group and 60.8 percent patients of placebo group experienced at least one adverse event. Of these, 3.3 percent of Synvisc-One group and 1.5 percent of placebo group had adverse events that were assessed by the investigator to be related to the study treatment.

Adverse events in the target knee occurring in more than one patient in either group-safety population are summarized in the table. Treatment-emergent adverse event rate of the two groups are comparable to each other.

Safety and effectiveness. The Panel will be asked a question about the overall safety and efficacy of this device.

Key efficacy results. I will present, primarily, the result of the applicant's initial submission. Dr. Lao, a FDA statistician, will present the applicant's analysis requested by FDA and FDA's analyses of the primary and secondary endpoints. The following slides will demonstrate that, for the primary endpoint, there was a

statistically significant difference in the least square mean of WOMAC A scale, using analysis of covariance.

The clinical significance of this change will be a question for the Panel. There is also a question to the Panel. How much mean difference between the two groups should exist in order to be clinically meaningful? There are a number of evaluations of secondary endpoint that were variable in their result. The Panel will be asked the question on the result of a secondary endpoint by the various methods.

Primary endpoint. This is the analysis of primary endpoint submitted to FDA in the original supplement. There was a statistically significant difference in the primary endpoint between the two groups in favor of Synvisc-One. The difference in the least square mean change from baseline between the two groups was 0.15 out of a five Likert scale. It will be a Panel question whether such a difference of 0.15 out of five scale is a clinically meaningful difference between the two groups.

Summary of results. The least square mean difference from the baseline through the 26 weeks on the WOMAC A scale, between the two groups, was 0.15

on a five-point scale. The primary endpoint has a p-value of 0.047. The Panel will be asked the question about the effectiveness of device, based primarily on these two findings.

The applicant's predetermined secondary efficacy end point. One of the secondary endpoints was to analyze the difference in WOMAC A sub-score from baseline to week 26 assessment between the two groups. The secondary endpoints were to analyze the difference in the WOMAC A1, WOMAC C, PTGA and the COGA subscores over 26 weeks and from baseline to week 26 assessment between the two groups. Another secondary endpoint was responder analysis according to responder criteria of OMERACT-OARSI set between the two groups. There was no pre-specified adjustment for the Type I error. The secondary measures were described by the applicant.

WOMAC A at 26 weeks. Analysis of WOMAC A at 26 weeks shows no statistically significant difference between the two groups, according to an analysis of covariance.

Categorical analysis of secondary endpoints. The applicant submitted results of analysis of the above ordinal data, using proportional odds model, cumulative logit model for

PTGA, COGA and WOMAC A1. There were statistically significant p-values in PTGA, COGA and WOMAC A1. Dr. Lao will discuss the applicant's and FDA's proportional odds analysis of PTGA, COGA and WOMAC A1 in his statistical presentation.

This is the applicant's proportional odds analysis for COGA. There were statistically significant differences in week 26 and overall 26 weeks.

This is the applicant's proportional odds analysis for WOMAC A1. There were statistically significant differences in week 26 and overall 26 weeks.

The applicant's responder analysis according to OMERACT-OARSI criteria. The proportions in the responder rate between the two groups at 26 weeks and over 26 weeks were analyzed. There were no statistically significant differences in the proportion in the responder rates between the two groups, either at 26 weeks or overall 26 weeks. Dr. Lao will discuss statistical issues regarding the analysis of primary and secondary endpoint in his statistical presentation.

Repeat treatment phase of the study. After the completion of safety and effectiveness assessment

at the week 26 visit, patients were offered participation in repeat treatment phase of the study, which lasted for an additional four weeks. Study was conducted to monitor only safety after the initial 26-week study.

The same knees of the patient, as were treated in the initial treatment, were injected with a second injection of the same doses. It is an observational study. There were 77 patients for Synvisc-One Synvisc-One treatment group, and 83 patients for placebo Synvisc-One treatment group. Adverse event rates of both groups were comparable to each other. The adverse event rate of Synvisc-One Synvisc-One treatment group were less than the adverse event rate of the placebo Synvisc-One treatment group. As to FDA's statistical presentation, Dr. Lao from FDA will discuss in detail both applicant's analysis requested by FDA and FDA's analysis in his presentation. Thank you.

DR. LAO: Good morning, Panel Chairman, Panel members, ladies and gentlemen. I appreciate the opportunity to speak to statistical perspective for PMA 940015, Supplement Number 12. My name is Chang Lao from FDA Division of Biostatistics, Office of Surveillance and Biometrics.

This is the outline for my talk and the -- of statistical component of the submission. First I'll talk of sample size, statistical models on repeated measures, effective result by FDA, and a summary.

Sample size determination is superiority trial with the primary endpoint based on the main difference in the change from baseline of WOMAC A pain score between the two groups. Applicant's assumption of the sample size calculation is based on two-sided Type I error rate, which -- rate equals five percent, power rate, 80 percent. There's the probability of finding a true secondary difference between the two groups, equals 80 percent. Overall treatment difference based on mean change from baseline, 0.297. Common standard deviation over two groups, 0.725. Effect size is a ratio of the difference divided by the standard deviation, equals 0.41. This is close to the median effect size, .5. The expected dropout rate, 25 percent.

So sample size in each group, 93 subjects per group, unadjusted for 25 percent dropouts equals 124 subjects per group, and adjusted for 25 percent dropout. And the footnote, sample size calculation was based on t-test, not based on the repeated

measures.

This is the total sample size distribution by country and site. As you can see, there are six countries, a total of 21 centers. The number of subjects between the two groups appear to be very close from each country, from country to country. It appears that a one-to-one randomization worked pretty reasonably here based on this, Table 1.

Statistical models. Because I concentrate on statistic inference, I needed to spend some time talking of what kind of model we used and plus some statistical terminology. We used the mixed model on the repeated measures. Applicant model mean change from baseline over 26 weeks on treatment, site, visit, treatment-by-visit interaction, and the baseline WOMAC A score. They include site here. FDA model mean pain score not changed from baseline, mean pain score over 26 weeks, with similar covariant but without site. We chose site as a random effect.

So FDA analysis of covariance, ANCOVA, tests null hypothesis, no difference in overall least square means, LSMEANS, in WOMAC A pain score and other WOMAC scores averaged over 26 weeks between the two groups. The reason we call it least square mean is because the model -- least square mean.

Okay. Now, I'll spend time talking of criteria for the model selection. Purpose: Find a model which fits best to the observed WOMAC A data by jointly modeling mean and variance-covariance structure. Second: Find a better variance-covariance measure structure among repeated measure of visits in terms of residual maximum likelihood, for example, auto-regressive of order one correlation. AR(1) correlation assume various -- that the covariance decreased as time advances expiration rate.

And we used the likelihood ratio test for comparing full and reduced nested model. Also, we used the Akaike information criteria, AIC, or likelihood ratio, for selection of different models under the same data set. The model should be sufficiently complex to fit the data best, but also a parsimonious model. A simple model, if possible.

Applicant's original analyses, ANCOVA. ANCOVA on change from baseline, CFB, site is fixed effect. So each site, the same effect, same variance. So their model, CFB_{ij}, i, subject, j, visit, equal the linear combination of this covariant -- $_{ij}$, for subject i, visit j.

Statistical models on repeated measures.

FDA model: Repeated measure analysis of covariance. We model mean pain score, Y_{ij}, over 26 weeks of the patient covariant. So the -- of Y_{ij} -- mean of the Y_{ij} is the linear combination of this covariant, and the Beta is intercept. E_{ij} is the error term. The above parameters, Beta 1 up to Beta 5, estimate by generalized least square, is the SAS software -- mixed software.

Different questions answered by change from baseline versus FDA's ANCOVA on mean. Change from baseline, applicant used -- either used in the randomized or observational study. The question to ask, are the profiles of the average change over all visits equal between the two groups? But ANCOVA on mean FDA used, appropriate only for randomized trial. This is randomized trial, anyway. So the question we asked, what's the expected true treatment effect on means over all visits, given that each subject has the same baseline value? We assume the population distributions of baseline values are equal between -- equal randomization.

Comparison between the two different models, applicant versus FDA. The ANCOVA model on means over 26 weeks, FDA, always has a smaller variance of treatment difference. It means more

efficient or more powerful than the mean change from baseline model, except when the correlation between repeated visits reaches 1.0, perfect correlation. That's very rare.

And the relative efficiency, which is the variance of ANCOVA based on mean and the variance changed from baseline, can be simplified, one plus -- divided by two -- is the correlation coefficient among repeated visits. Assume this is a compound symmetry correlation. And post-treatment visits equal one -- otherwise, general case, the efficiency is dependent on the number of repeated visits and the correlation structure.

With the FDA model, the treatment effect, averaged over all 26 weeks and at each visit, is measured by the difference in the estimated least square mean between the two groups and more likely is a more powerful approach. However, no matter what the model, we are testing the null hypothesis of zero difference. Because this is a superiority trial, this does not guarantee a clinically meaningful difference, which may be not zero.

Table 2 is the FDA model and the difference between the mean, least square mean, minus 0.15, the same -- changed from baseline. But a standard error

of the difference is 0.072, which is smaller than this -- standard of 0.076. So the 95 confidence intervals for the difference changed from -- changed from baseline, a change of p-value from 0.047, changed from baseline model into the 0.032, into the baseline model on mean. This is the primary endpoint on WOMAC A, on the auto-regressive correlation.

This is the FDA's analysis of primary endpoint of observed and fitted mean WOMAC A score on repeated measures. To answer Dr. Blumenstein's question, this table will give you the baseline, four week, eight week, up to 26 weeks. Sample size and -- at the beginning, 124 for the Synvisc, and 119 and up to 115, the 26th week. So similar interpretation for the control group. We have observed mean of 1.45 for week four, and fitted by the model, 1.48. The difference between the two is 0.0 -- a negative 0.03. So this is the residual for Synvisc, O minus F, observed model fitted. If you look down this column here, the difference up to the second decimal point. So overall, the model fitting, I would say, pretty reasonable. Similar situation for the control group.

And the missing data here at the beginning, 124, at the end of study, 115, so nine patient visits and about five or six percent. So the percent of

missing data not too severe here. And also, in the mixed model, we assumed those missing data and the -- so it means the probability of missing data is independent for the future observed data, so that assumption used -- mixed.

Table 4 is FDA ANCOVA for secondary endpoint on mean results. This one is -- the ordered result here is the secondary endpoint, based on -- covariance, repeated measure on least square mean. And the least square mean for the WOMAC A1, walking pain, at 1.44 for Synvisc-One, and placebo, 1.63, a difference here of standard error and 95 confidence interval and the p-value. And the confidence interval will give some clinical interpretation, and the p-value only gives you probability statement. So the PTGA, COGA and the WOMAC C and -- significant, except WOMAC A1 based on a mixed model.

Secondary effective endpoint continued, WOMAC A1, WOMAC C, PTGA, COGA, and OMERACT-OARSI. In the original submission, the applicant prepared a different approach for each endpoint. The first was a mixed model for change from baseline, for WOMAC C because WOMAC C has 17 questions there, and they used every -- 17 questions and each question has a zero to four, five point.

And that they also used a proportional odds model for the ordinal categorical data for A1, PTGA and COGA. That's only based on one question, each question for five points, zero to four. And the final rating was binary analysis for the OMERACT-OARSI. That's a responder/nonresponder rate, odds ratio equals 0.66 overall. P is not as significant for the binary analysis.

Secondary effective endpoint continued. At FDA's request, the applicant prepared a mixed model on change from baseline for WOMAC A1, PTGA and COGA. Only WOMAC A1 was statistically significant. P equals 0.029, based on mean change from baseline versus P equals 0.017 based on FDA's mean score over 26 weeks. So both are significant from zero.

For proportional odds model, the applicant provided graphical results for PTGA, COGA and WOMAC A1, by various cutoff point of the clinical outcome because we have a total of five outcomes, no pain, mild, moderate, severe, extreme, to show the validity of proportional odds model. The problem is it appears no existing computer software is available to test proportionality of parallelism assumption of slopes from different cutoff points of clinical outcome.

Comments on the applicant's generalized estimating equation model based on proportional odds assumption. If we let Y_j, for j as outcome, equal probability, the outcome less or equal to j, condition on vector of covariate X, which is jth cumulative response, Y_j, probability given a set of covariate X, group, site, visit, visit group interaction, and baseline.

By logistic regression for p covariate, we have a logit, Y_j, which is defined by -- of the probability, Y less or equal than j, divided by probability, Y greater than j. J goes from zero, one, two, three, four, for the no pain, mild, moderate, severe, extreme. As you can see from this logistic regression model, if x_1 is the treatment, ß is the regression coefficient for the treatment, assume proportional odds model, assume. It doesn't matter which cutoff we use, use G equals zero, or zero plus one, or zero plus one plus two versus as -- in the denominator. The $ß_1$ equals common ß. No change. So that's the proportional odds model, calculate odds ratio. And so odds ratio here, based on this logit model, is E to the ß power, $ß_1$ power -- $ß_1$ -- GEE output based on logistic regression model.

Question to ask, Are slopes parallel? Does $ß_1$ equal ß? Cutoff point j, does it matter? This is the number one question. Applicant's response: The data provide a graphical visual inspection of odds ratio, and that there are 95 confidence intervals from different cutoff points, j for zero, one up to four, shows overlapping of 95 confidence intervals. So they believe cutoff point does not matter. Problem: No formal hypothesis testing. Most 95 confidence intervals contain one for COGA, PTGA and the WOMAC A1, which we'll show in the next graph.

Figure 1 is the applicant's justification of proportional odds model for three different secondary endpoint. The first one is COGA and the next one is PTGA and the WOMAC A1. There's different cutoff points. This is zero versus one, two, three, four. Zero, one compared with two, three, four. So by combined five-by-two data into two-by-two table, a different comparison, you can see from this chart here, the point estimated odds ratio here, most of them are less -- 95 confidence intervals must include one. The point estimated odds ratio, most of them are less than one. But a 95 confidence intervals cover one. And they used this graph to show the proportional odds model is reasonable.

This is Table 5, the summary of statistical significance testing over 26 weeks. Here, the primary endpoint. Applicant's mixed model on change from baseline, site fixed. It showed the same effect, clinical effect. No variability from site. So the p rating is 0.047. And we assume each set is different clinical effect. Each site has -- a p equals 0.032 random effect model. So it actually improved the p rating from 0.047 to 0.032.

Secondary endpoint, no multiplicity adjustment and -- adjustment. FDA requested the applicant to also do the mixed model and for those outcomes, A1, PTGA, COGA -- except the A1. And for the PTGA, COGA and WOMAC C, none of them are statistically significant based on mixed model, assuming -- distribution versus the FDA mixed model. So this -- except that this, the proportional odds model repeated the measure on the generalized estimated equation model using the covariate, site, baseline, visit, visit by group interaction. You have significant p-value, A1, PTGA and COGA. For the binary responder, we agree, is not significant, p equals 0.059.

So summary. Primary WOMAC A pain score showed a difference Synvisc-One minus placebo equal

0.15. This is about three percent of the five-point scale. Applicant and FDA agree, statistically significant. The question is, is it clinically significant? The Panel question.

Secondary endpoint, WOMAC A1, ANCOVA. Treatment difference equals minus 0.19, Synvisc-One minus placebo. Applicant and FDA also agree, statistically significant and -- zero difference, but no multiplicity adjustment. PTGA, COGA, not significant by ANCOVA model by FDA, but they are significant by proportional odds model. But again, no multiplicity adjustment. That's another Panel question. Thank you.

Back up a slide. A different way to look at the odds ratio, based on two-by-two table, not based on model. If, Panel, you're interested, I can show you, otherwise I'll stop here. Thank you very much.

The next speaker will be Dr. Wang from FDA who's talking about post-approval study.

DR. WANG: Thank you, Dr. Chang. And good morning, distinguished Panel Chair and members and the welcomed guests. My name is Cunlin Wang. I'm an epidemiologist in the Office of Surveillance and Biometrics. I'll now present post-approval study

consideration for Synvisc-One device.

And first, please be reminded that discussion of the post-approval study, prior to a formal recommendation on approvability of this PMA, should not be interpreted to mean that we are suggesting the Panel find the device approvable. The plan to conduct the post-approval study does not decrease the threshold of evidence required to find the device approvable. The premarket data submitted to the Agency and discussed today must stand on its own in demonstrating a reasonable assurance of safety and effectiveness in order for the device to be found approvable.

The main objective of conducting post-approval studies is to evaluate the device performance and the potential device-related problems in the broader population over an extended of period of time after premarket establishment of reasonable device safety and effectiveness. Post-approval studies should not be used to evaluate unresolved issues from the premarket phase that are important to the initial establishment of device safety and effectiveness.

Generally, the reasons for conducting post-approval studies are to gather post-market

information, including the longer-term performance of the device, community performance, which is the device performance in a broader patient population treated by average physicians as opposed to highly selected patients treated by leading physicians in the clinical trials. Post-approval studies are also used to evaluate the effectiveness of the device utilization training programs and evaluation of the device performance in sub-groups of patients since clinical trials tend to have limited number of patients and may not include all sub-groups of general patient population. In addition, post-approval studies are also used to gather data on device real-world experience and to monitor device-associated adverse events, especially rare adverse events that were not observed in the clinical trials. Finally, post-approval studies also enable issues and concerns raised by the Panel members to be addressed.

Currently, the Sponsor did not consider a post-approval study is necessary and therefore did not provide a post-approval study plan. They identified a few issues that may be considered in assessing the need for a post-approval study of Synvisc-One in the United States. First, the clinical study supporting this PMA supplement was

solely conducted in Europe, and literature has shown that patients' characteristics may be associated with the treatment effects of the device.

Second, the follow-up of this PMA study was 26 weeks for initial phase and four additional weeks for repeat phase, while intra-articular injection of similar devices has demonstrated the treatment effects extended to 12 months after injection.

And third, literature has also suggested that, compared to sodium hyaluronate, cross-linked hylan G-F 20 used by Synvisc may be associated with increased risk of severe acute inflammatory reaction; the exact mechanism and the long-term consequences remain unclear.

Given this consideration, if a device is recommended for approval at a later date, we would like the Panel to comment on the need to evaluate the device in U.S. population in a post-approval study. And if a post-approval study is recommended, we would also like the Panel to discuss the following items: objective, clinical endpoint, study size, comparison group, duration of follow-up, and other specific issues that you would like to be addressed. That's it. Thank you.

DR. MABREY: I'd like to thank the FDA

speakers for their presentations. Does anyone on the Panel at this point have brief clarifying questions now for the FDA before we get into our general discussions? You may also ask the FDA more in-depth questions during the Panel deliberations coming up. Any specific questions for the FDA to clarify their presentations? Dr. Goodman?

DR. GOODMAN: No.

DR. MABREY: Dr. Olsen?

DR. OLSEN: No, I have not.

DR. MABREY: Okay. Dr. Skinner?

DR. SKINNER: No.

DR. MABREY: Dr. Blumenstein?

DR. BLUMENSTEIN: No.

DR. MABREY: Thank you. Ms. Rue?

MS. RUE: No.

DR. MABREY: Ms. George?

MS. GEORGE: No.

DR. MABREY: And Dr. Evans? Okay.

MR. HALPIN: Dr. Mabrey?

DR. MABREY: Yes.

MR. HALPIN: I just wanted to point out that the Sponsor is ready to answer Dr. Blumenstein's clarifying question from earlier, if you're ready for us now.

DR. MABREY: Yes, this would be an appropriate time to clarify Dr. Blumenstein's question.

MR. HALPIN: Okay, I'd like to have Dr. Nancy Silliman --

DR. MABREY: I'm sorry, to present an answer to Dr. Blumenstein's question.

DR. SILLIMAN: Thank you. My name is Nancy Silliman. I'm a Vice President of biostats and stat programming at Genzyme. Slide on, please.

So first I would like to go through the reasons for discontinuation, and overall, the dropout rate was relatively low; it was eight percent. In the Synvisc group there was one patient who dropped out for an adverse experience, one patient was noncompliant, one who wished to withdraw, and six who dropped for lack of efficacy.

In the placebo arm, there were three who dropped for adverse experiences, two who were noncompliant, one who wished to withdraw, four who dropped for lack of efficacy, and two who dropped for other reasons. Slide on.

This shows the amount of patients, patient data available at each visit for the Synvisc arm as well as the placebo arm. And you can see, for

Synvisc, it baselined. There were 124 patients, and there were a high number of patients through week 8, maintaining a still high number through week 26. And a similar distribution was seen in the placebo arm. Slide on.

And this is just a little bit more detail. We did do quite a bit of sensitivity analyses around the missing data. The column in blue was our primary endpoint, with no imputation of missing data. So I would just like to clarify. Since it was a repeated measures analysis, we did actually use all available visit information for each patient, so patients that dropped out would just contribute less information to the overall estimate of the treatment effect.

We looked at worse case analysis, which is this second row here, where we assumed that Synvisc-One patients showed no change from baseline after withdrawing, whereas the control patients showed their best results observed after withdrawing. The treatment effect was similar. The p-value was .069. We also looked at baseline carried forward where, for all patients who dropped out, we carried forward their baseline value. This is this column. The treatment estimate again was similar. The p-value was .04. We looked at a mixed baseline observation

carried forward, the last observation carried forward analysis, which is something commonly seen in drug studies of pain. That's this column here.

And this one is of interest because here we assume that, for patients who withdrew due to an adverse event or lack of efficacy, we're carrying forward their baseline values, so that's conservative, assuming there was no treatment effect. And then we used last observation carried forward for all the other patients.

And then, finally, we did a best case analysis, which is the opposite of the worse case. So now we're assuming that control patients -- we carried forward their baseline after they withdraw, and Synvisc-One patients, we carried forward their best observed value. So you can see the treatment effects are consistent and the p-values are also all fairly consistent.

DR. BLUMENSTEIN: I don't know whether -- I had a couple more questions that kind of bear on this issue, but it's getting into the weeds. So should I do it?

DR. MABREY: I think we can start to move into our general Panel discussions, and I'd be more than happy to start with you, Dr. Blumenstein.

DR. BLUMENSTEIN: All right. On Page 44 of your SAP, you have a little snippet of SAS code there that you say was used to do the primary analysis, and there's an element of that SAS code that I don't understand, and it might have bearing on this.

DR. SILLIMAN: Slide on. This is the actual SAS code that we used.

DR. BLUMENSTEIN: Okay, it's the second line. Could you explain --

DR. SILLIMAN: Yes.

DR. BLUMENSTEIN: -- what that means?

DR. SILLIMAN: Yes, absolutely. So let's just make sure that we're looking at post-baseline visits. Visit one was baseline, visit two was week -- no, I'm sorry. Visit one was screening, visit two was baseline, visit three was week one, and we didn't collect any efficacy information. Visit four was week four. So this is just making sure that we're using post-treatment efficacy.

DR. BLUMENSTEIN: I didn't know what the visit numbers meant.

DR. SILLIMAN: Yeah, I apologize for that.

DR. BLUMENSTEIN: Okay, thank you. For the rest of you, it's okay.

DR. MABREY: Thank you for clarifying that.

At this point we'll begin the Panel discussion portion of the meeting, as we already have. And although this portion of the meeting is open to public observers, public attendees may not participate except at the specific request of the Panel.

I'll just keep going. Dr. Blumenstein, do you have any questions for either the FDA or the Sponsor? And I'll remind the Panel that this is often a good time, if you have in-depth questions, to give both the FDA and the Sponsor a heads-up so that, over lunch, they can provide a more in-depth response. And so for both the Sponsor and FDA, if you think your answer is going to take more than a couple of minutes and you'd like some time to work on it, just say so and we'll expect your answer in the afternoon.

DR. BLUMENSTEIN: Okay. So my next issue has to do with the lack of control of Type I error over this secondary endpoint, and the Sponsor has stated that they're not making a claim and therefore it's not relevant. But nonetheless, we're being presented with an array of analyses based on the secondary endpoints, and we are subject, like it or not, to the possibility of coming to false positive

conclusions, especially since there's probably correlations between these endpoints. But I just wanted to ask, there was something in the FDA briefing document, and then that said that the Sponsor had not pre-specified a hierarchy of testing of the secondary endpoints.

But yet I found a statement in the SAP that says, the OMERACT-OARSI responder criteria is therefore considered to be the most important secondary efficacy endpoint in the study. Yet I don't see that declaration in the SAP carried forward in the presentation. Would you care to comment?

DR. SILLIMAN: Sure. So we didn't mean that statement to imply that we were going to do any sort of formal testing, looking at the OMERACT-OARSI first. Slide on. So the FDA had asked us, after the study was finished, to come up with a method of adjusting for Type I error over the secondary endpoints. We had proposed a hierarchical sequential testing order, and you see here the OMERACT-OARSI responder analysis was actually marginally significant. Over the 26 weeks the p-value was .059.

DR. BLUMENSTEIN: But this doesn't represent what you said in the SAP?

DR. SILLIMAN: Right, that was our best

thinking on the subject at that time. But again, in the SAP, we specifically said we weren't planning to adjust for multiplicity for the secondary endpoints.

DR. BLUMENSTEIN: So when you submitted this list of -- this hierarchy of secondary endpoints to the FDA, this was after you had already analyzed the data?

DR. SILLIMAN: Yes, absolutely. And we noted that in our response. The FDA asked us this question after we had submitted our PMA document, so we clarified that it was post hoc.

DR. BLUMENSTEIN: I'm done for now.

DR. MABREY: Okay. Ms. Rue, questions for the FDA or the Sponsor?

MS. RUE: I don't have any questions at this point.

DR. MABREY: Okay. Ms. George?

MS. GEORGE: I guess I just have one question right at the moment, and maybe it's because I'm confused about all the statistic stuff. But I know, in the Genzyme package they presented, it looks like seven different analyses of data, the one that they submitted initially, and then there's six more that the FDA either did themselves or requested. So I guess what I'm trying to understand is, is which

technique is the FDA actually asking of the Sponsor to focus on for us to be able to make the decisions as to whether they met or did not meet their endpoints.

DR. LAO: This is Chang Lao. We tried different models because we have found the best model fitted data best. So in terms of correlation of various covariance measures, we tried -- at the beginning we tried like an unstructured -- and a compound symmetry, and finally -- and we agree with the Sponsor. The last covariance measure is the best to fit the data best.

So, finally, the only difference between the Sponsor's model and the FDA's final model is they choose site as a fixed effect, which assumes each site has same variance, same clinical effect, and have some sites as a random effect, and last, some variability among different sites. Also some slide today, different mean response from site to site. That's the only difference between the two different models.

And finally, we compared observed and fitted model, and we feel the analysis of covariance on mean, least square, fitted data pretty well. That's about a history of model fitting procedure.

MS. GEORGE: Okay.

DR. MABREY: Does that answer it? Other questions, Ms. George?

MS. GEORGE: Not right now, no.

DR. MABREY: Dr. Evans?

DR. EVANS: Yes, I have several questions and they're sort of spread around, so maybe I can sort of fire them off and you can respond to them after lunch. But first allow me to thank the folks at Genzyme and the FDA for their comprehensive efforts. I appreciate the complexity of the issues of clinical relevance and conducting pain trials, and I think you've done a nice job trying to understand the data.

So question number one is -- some of them are just clarifications, and others are a little bit more in depth. The first question I have, this was -- your pivotal trial was a blinded trial, and I was wondering if there was any assessment of the success of the blinding in particular because, you know, pain is a very subjective measure and because of the subjective endpoints, I think it's important to get an idea about how successful the blinding was. So that's question number one.

Question number two is actually more of a

comment or a question for Dr. Dworkin, who -- I found your presentation very informative and particularly the distinction between clinically meaningful group difference in contrast to clinically meaningful changes in individual patients. And you provided a nice list of considerations for defining what would be a clinically meaningful group difference. And so not to put you on the spot, but I was wondering if you had an opinion about, given the characteristics of this trial and this syndrome, what you thought a clinically meaningful group difference would be in this particular case.

Question number three or clarification number three was just a terminology issue. I know, in your presentation, you talked about treatment effects, but then you talked about effect sizes, and I often use the term interchangeably, but you had distinct definitions for those, so I'd just like to clarify what was meant by that.

The fourth question was about the design of the trial, and you did a nice job explaining how you sized the trial with sample size and power calculations, and you stated that you selected an effect size of .297.

And oftentimes, when we size these trials,

we pick a minimum clinically relevant difference, and I was wondering whether that was selected because that's what you believed this was. But I think you used terminology that the .297 was an estimated treatment effect and not necessarily a minimum clinically relevant difference, and I would just like you to comment on that if you could.

And my last question was -- actually, it's sort of directed at both Genzyme and the FDA folks, and the FDA presentation, I thought, brought about a very important issue here, and that is about one thing that was -- that came across somewhat nicely is actually because of the different models that were fit, you actually have conducted sensitivity analyses of sorts in looking at the consistency of at least the qualitative interpretation as you vary different models.

And I would like to ask both the FDA and the Genzyme folks to comment on, one, as you fit these models, what did you check in terms of model assumptions? All models have some sort of assumptions associated with them and I know, in the FDA presentation, I think you mentioned that there's actually no software that evaluates some of these assumptions.

And so I think sensitivity analyses are the key, and I think checking model assumptions is also very important and what were the results of that. And as an extension of that, I hate to be one who suggests an alterative analysis, but given that you've probably analyzed this more than you care to -- but there are methods that are "model-free or more robust" to assumptions, in particular nonparametric things that don't require assumptions about distributions and things like that, and other types of methods, Way and Johnson (ph.) type of things that are essential model-free.

And since you don't have modeling, you don't have the assumptions associated with the modeling, and there are some methods that could be explored that essentially eliminate the problems with assumptions or having to make them. And so I just sort of throw that out as an idea and whether that's been tried.

MR. HALPIN: So I think we can respond to some of the clarifying questions right now, if that would be appropriate.

DR. MABREY: That would be appropriate, yes.

MR. HALPIN: Okay, great. First I'd like

to have Dr. Polisson come up and speak to the treatment effect and the protocol of 0.297, and also touch base on comments about OMERACT-OARSI as the most important secondary endpoint.

DR. SILLIMAN: All right. So I'm going to speak to the choice of the .297 for the power calculations. First let me clarify, when we sized the trial -- and I'm sorry for the confusion. So actually, maybe, let me back up and start with the way we define effect size here is the treatment difference divided by the standard deviation in the control group. So I'll try to be very clear about whether I'm talking about a treatment difference or an effect size. The .297 was actually an observed treatment difference in a previous open-label Synvisc trial versus steroids.

So the -- all right. Let's see, can I have the slide on? So this is just a recap of the power calculations and that the estimate of the treatment difference of .297 was based on this Kayborn (ph.) study.

Slide on. This was an open-label study of Synvisc versus Arristaspam (ph.) and designed very similarly with WOMAC A and almost exactly the same treatment schedule. There was no visit 18. Sorry,

no week 18 visit, so we just interpolated responses in that study and then we -- for each patient, we averaged the overall mean change from baseline across the study visits, post-treatment study visits, and we used this to come up with our estimate of .297 for the treatment difference as well as -- I think it was .725 for the standard deviation. Slide on.

One important point about the Kayborn study was that this was an open-label trial and so treatment effects tend to be much larger in the open-label study. We postulate this could be one reason why the observed effect in the current study is less than .297.

So, you know, power I show you design the study that's related to the Type II error, the risk of not being able to observe a significant difference in your study. Once you're done with the study, then you're looking at the p-value for the primary endpoint. You're interested in preserving the Type I error, for example, at five percent.

So, again, this .297 was chosen, based on the Kayborn study, as sort of our best estimate of what we might see as a target treatment effect. It was not chosen to be any sort of a minimum difference that would be considered clinically meaningful. So

you could certainly have a treatment effect less than .297 and still consider it clinically meaningful.

DR. EVANS: Although, just to clarify -- so if that was the case, then would the current trial actually have been underpowered for effect sizes smaller than .297?

DR. SILLIMAN: Well, we also estimated a dropout rate of 25 percent, and we saw about eight percent, so I think that the dropout rate and the treatment effect, treatment difference, kind of wash each other out, so that we -- you know, we believe that the study was adequately powered. I might also try to answer the question about the blinding.

We did not assess, at the end of the study, whether the patients guessed which treatment they received. However, the injection adverse event rates were similar between the arms, suggesting that that wasn't a cause for un-blinding, as well as we used a blinded injector so that they were the ones communicating with the patient.

MR. HALPIN: Okay, I'd like to have Dr. Stephen Lake come up.

DR. LAKE: Hi, my name's Steve Lake, Genzyme biostatistics. Slide on, please. And so we actually -- at the time we responded to FDA, we did

not have a formal test of the proportional odds assumption, but then we subsequently actually did identify two tests for that proportional odds assumption.

So just to refresh what Dr. Lao said, the proportional odds model, when we have ordinal data as a commonly used extension of logistic regression for ordinal response variables, what we do is we model the cumulative logits, and there are assumptions with this proportional odds model, namely, that it assumes that the odds ratios associated with covariance, such as treatment effects, are the same, regardless of which cumulative logit is used. So if you're looking at the treatment effect -- and this is what Dr. Lao presented.

He showed a graphical representation of the odds ratios across the cumulative logits. We can actually test that assumption of whether or not those are equal or show large deviations from the assumption of proportional odds. So the next slide, please.

DR. EVANS: Could I just clarify? So the null of the test is that there's no violation, is that right?

DR. LAKE: Yes. So the null is that there

is no violation, yes. So this is the actual table here of p-values and a p-value less than .05 would indicate evidence against proportional odds. And you can see that there is test for the GEE proportional odds assumption, and that's the test of overall on the first row there, and you can see that this is what Ms. Elkins indicated, that the p-values are all greater than .05, indicating that the proportional odds assumption does hold.

And then we also looked at each specific post-baseline proportional odds and tested that assumption as well. And there was only one out of 15 that indicated that, at week 12, in that COGA, that there was a deviation for proportional odds. So we feel comfortable that use of the proportional odds model to analyze ordinal data is warranted in this situation.

DR. EVANS: So let me just clarify, I guess, the way I would state it. So these tests were not significant, essentially stating that you looked for violations to the model assumptions and did not find them?

DR. LAKE: That's correct.

DR. EVANS: Which is good, although just to clarify, is distinct from confirming that the

assumptions indeed hold. So, in other words, you failed to find evidence against it, but it doesn't mean you found evidence to support it.

DR. LAKE: That's correct, yeah. And I think the graph that Dr. Lao presented showed that, you know, there aren't no large violations, visually, as well.

MR. HALPIN: I'd like to have Professor Chevalier, who was a clinical investigator in our study, just come up and speak briefly to blinding.

PR. CHEVALIER: I am Xavier Chevalier, head of the department of rheumatology in Paris, and I was one of the senior investigators in this trial. And I have my travel taken charge by Genzyme and sometimes there are fees as a consultant for Genzyme. I would like to answer on your question on the blinding, which is very important for this trial, of course, for a patient.

The surveying was completely hidden, so the patient couldn't know whether he received the placebo or he received the drug. And in this kind of trial is an investigator who -- the patient, who was not, of course, the one who injected the product. So taking together, the patient couldn't know whether he

receive or not the -- or the placebo.

MR. HALPIN: Okay, I'd like to have Dr. Dworkin come up briefly and answer the question.

DR. DWORKIN: That is a great question, Dr. Evans, so let me preface it by saying that most of the research I've done for the last 20 years has involved drugs, not devices, and so we've done a lot of studies of anti-depressant medications, anti-convulsive medications, in neuropathic pain conditions, like diabetic neuropathy, but also more recently in low back pain and in osteoarthritis.

And so my perspective on your question really comes from that drug and particularly anti-depressant and anti-epileptic background. In that arena, a delta of active treatment versus placebo because we do have inert placebos, obviously, in drug studies, of 1.0 out of 10 is a common delta. We do find, you know, between group differences of one or even a bit more out of 10, out of 0 to 10 pain scale. Of course, we have to remember this is a five-point scale in this pivotal trial.

And so I think we wouldn't be here today, is my guess, if the delta in the pivotal Synvisc trial was of that magnitude because I think that seems pretty obviously clinically significant, and

the clinically significant group difference would be meaningful. But of course those drugs, where the delta can be 1.0 out of 10, are drugs that cause nausea and constipation and serious cardiac toxicity in some cases.

Anti-epileptics are associated with Stevens-Johnson syndrome. And that's all taken into account in tolerating between group differences that can be one out of 10, or even one and a half out of 10, almost never more than that. But of course, here, the between group difference was less and I think we wouldn't be here today and -- you know, I hadn't thought about it until your question. I don't think we'd be here today if Synvisc-One was associated with an elevated rate of Stevens-Johnson syndrome or cardiac toxicity or the development of addiction, but I don't know what the addiction to Synvisc-One would be, as we deal with, all the time, in the drug world.

And so I think, from the perspective of a treatment benefit, that is clearly modest, you know, but that has a 26-week duration with what seems to me a very, very low rate of adverse events, and you have one injection that gives benefit for 26 weeks. I think, from my perspective, that's clinically

meaningful in the context of, you know, my background in drug development. One of the things I do but I didn't mention is publish consensus treatment guidelines for neuropathic pain.

And so when I view this in terms of the benefit versus risk tradeoff that we obsess about when we publish consensus treatment guidelines, to me it seems a clinically meaningful benefit. You know, 26 weeks after a single injection, with no adverse events. And I'd be happy to answer a follow-up question, if you have one.

MR. HALPIN: In regard to your last question, Dr. Evans, I think we're probably going to elect to answer that after lunch. I was wondering if you could restate it. Oh, you have it? Okay, we've got it. Thank you.

DR. MABREY: Dr. Goodman.

DR. GOODMAN: I'd first like to thank the Sponsor and the FDA for their excellent presentations and for the opportunity to comment on this submission. I have a number of very basic questions, and I'm just going to read them off and you can either answer them now or perhaps answer them later.

First of all, I'd like to know, from the Sponsor, how they think Synvisc actually works and

how it decreases pain. I didn't see anything in the submission to this effect.

Number two is I'm wondering when the product was introduced originally. It was originally proposed to have three injections and not one, and I'm wondering now, at this point, why they're going to a single injection rather than three injections, other than the reason that was given by the clinician at the beginning of this meeting. And I'm wondering why six cc's are being introduced and not two cc's, if they have any information on this.

Third, in terms of the inclusion and exclusion criteria, I'm wondering if they could explain the inclusion criteria for the K-L assessment. They ruled out, I believe, severe degenerative arthritis, but I'm wondering how they assessed and what groups they included with regards to the mild and moderate arthritis.

They also have excluded people with significant varus and valgus deformity, and I'm wondering how they defined severe varus and valgus deformity. The majority of our patients with degenerative arthritis have a varus deformity about 10 to 1 versus valgus deformity, and I'm wondering how they excluded people with varus deformity of a

severe nature or a valgus deformity of a severe nature.

I also would like to know if they have any data on rescue medication. Did the placebo group take more rescue medication than the treatment group? I'm wondering if they have also a control group where the patients came to visit the doctor and didn't receive any injections at all, either in the past, from their past data, or if they have another group as well.

One other point that I neglected to mention was with regards to inclusion and exclusion criteria. Other than the radiographic designation and the varus and valgus deformity, one of the exclusions was a tense effusion. So if patients come with mild to moderate arthritis, they generally have an effusion. At least probably half of them do. And I'm wondering how they excluded patients with a tense effusion. Did that make them enter into the severe arthritis group, or how exactly clinically did the people involved deal with patients who had a tense effusion?

Finally, the improvements were very modest, and even as in comparison, NSAIDs and other treatments seem to really have a very modest improvement. And I think the Sponsor gave a fairly

compelling reason how their single treatment fits into this paradigm, and I'm wondering, do they have plans to perhaps repeat the six cc injection in the future, if a patient might not respond the first time, or is the six cc injection going to be the be-all and end-all? Thank you.

MR. HALPIN: Okay, thank you. I think we can probably answer some of these questions now, and we may need to answer some of them after lunch.

DR. MABREY: Okay.

DR. HOLMDAHL: Thank you for all the questions. I would like to answer first how we got to the single injection and why we choose six cc's. That was actually based on the results of a pilot trial -- slide on, please -- where we evaluated various different combinations of volumes and number of injections and we rated, as I briefly mentioned in my presentation, the performance, both in terms of efficacy and safety of these various treatments. And the three times two mL here is the currently approved treatment, which was on the WOMAC A1, your results here, and this is the rank of the PTGA, and this is the rank of the COGA. And as you can see here, the one-time six mL performed at least as good as that, whereas the other various combinations here did not

perform as well as -- at least in our minds, as the three times two mL treatment. Next slide on, please.

And the reason why we did this to begin with was that we -- as I briefly also mentioned, that we had received requests, if it was possible to simplify the treatment, since the patients have to return a couple of times to get their full treatment. That is basically the justification for why we did this. And we knew that physicians were experimenting with simplification and alternative doses, so we thought it was the responsible thing to do, to investigate this. Yeah, can I have the slide on rescue medication from the -- yes, slide on, please.

So this is the rescue medication, the average mean daily use of paracetamol. So it's specifically to the rescue medication for the duration of the trial, as you can see here. And the Synvisc-One is shown in blue and the control is shown in red. And after about a month, the two curves began to separate, and then there is a trend towards greater average daily use of rescue medication in the Synvisc-One arm, although this difference, over time here, did not reach statistical significance. The p-value was 0.095. And then we will come back with the rest of the answers.

MR. HALPIN: We have someone who -- Dr. Murray is going to come up and address the mechanism of action question.

DR. MURRAY: Good morning. I'm Dr. Christopher Murray, a Senior Director in the Medical Affairs Group at Genzyme Biosurgery, an analgesic pharmacologist by training. Slide on, please.

When considering the mechanism of action of viscosupplements, the original hypothesis from a number of years ago that was first tested in racehorses was that when you have osteoarthritis, there's an observed degradation in the physical properties of synovial fluid inside the joint space. Slide on, please.

As you can see on this slide, the elasticity and viscosity and average molecular weight of hyaluronic acid inside the joint space for normal patients or normal volunteers, and you can see that when you get into a degenerative joint disease situation, that those physical properties get degraded. It was thought by the originators of viscosupplementation that if you replaced the degraded synovial fluid with a prosthetic device that had physical properties resembling normal synovial

fluid, that that would enable the joint to re-reach homeostasis, and because of that, that pain would be relieved and other symptoms would improve.

That hypothesis -- slide on, please -- has recently been tested in a human clinical trial that was published a couple years ago by some investigators in Australia. In that study, they took patients with relatively early staged disease, about 60 of them, and studied them for about six months after treatment with three injections of Synvisc. They took synovial fluid samples before the treatment and three and six months after the treatment. Slide on, please.

And the results of that study had the following findings: first, at month three, there was a statistically significant increase in the mean concentration of hyaluronic acid in the patient's joints, and the complex sheer module, which is a combination of elasticity and viscosity, increased significantly. There were similar effects at month six, although they did not reach statistical significance. So this was the first human demonstration of the proof of the hypothesis behind viscosupplementation.

Did that address your question,

Dr. Goodman?

DR. GOODMAN: Yes, but if you go back to your last slide, I think the controls you used -- do you want to put that back on, please?

DR. MURRAY: Thank you.

DR. GOODMAN: To the slide before that.

DR. MURRAY: RD-42, please. Slide on.

DR. GOODMAN: The one before that, please. So it doesn't appear that your normals are age-matched.

DR. MURRAY: That's correct, these papers were published at different times by different groups of authors, yes.

DR. GOODMAN: Do you have any idea what the properties would be on each match control?

DR. MURRAY: There were some studies that were done later on, with smaller numbers of patients that were studied in this particular paper, and they're about in the range of what you're seeing with that particular slide on osteoarthritic conditions. I don't have those data in a slide for you today, though.

DR. GOODMAN: So just to paraphrase, the age-matched controls had the same physical properties in their synovial fluid as the normals, age 18 to 27?

DR. MURRAY: There is a slight decrease with aging in those physical properties; however, they do not reach nearly the extent of the degradation that's observed with osteoarthritis.

DR. GOODMAN: Thank you.

DR. POLISSON: Let me see if I can answer two questions raised by Dr. Goodman. I don't have slides for this, so I'll just speak to them. It had to do with your question about a third control, I believe, and was there another control that was not treated by intra-articular saline or Synvisc-One, and the answer is no, and that was not part of the construct of this clinical trial, although it's an excellent question and one would love to have that information, but we did not do that as part of this program.

I believe your next question had to do with Synvisc-One being the whole enchilada, if you will, and I think there is -- if this is approved and used in practice, I think it's -- we should leave it up to the physician and the patient to decide which product would be most useful in that particular situation. I will say, however, that we did study Synvisc-One in a repeat phase, in an attempt to get some short-term adverse event data in case, you know, we did want to

go forward with using this much as you might do a steroid injection, and the safety looked pretty good. So again, I don't think that, you know, we think that this is going to be it. Both products will be out there and available.

DR. HOLMDAHL: Can I have the rescue medication slide again? I was advised that I misspoke to that slide, so I would like to show that again, the rescue medication slide. Slide on, please. So there was a less -- lower average daily consumption of rescue medication in the Synvisc arm. Thank you.

MR. HALPIN: And I think we would like to come back after lunch to answer your question specifically about the three different issues regarding inclusion and exclusion criteria.

DR. MABREY: Thank you. Dr. Olsen.

DR. OLSEN: I have one. It's actually more of a concern than a question. It reflects on the baseline characteristics or the population characteristics of these individuals who are enrolled, and I think reflects that they were European rather than studied in the United States, and that is that their mean BMI was 29. I'm the only rheumatologist here, but in my clinical practice I

see many -- when I see a BMI of 29, I actual notice it because so many of my patients have BMIs greater than 30 and even greater than 40. So one scenario for use in the United States, I would think, might be that such individuals who aren't ready to get their joints replaced because the orthopedic surgeons won't replace joints in such larger individuals, might be you need to lose weight, and there's ways of doing that now, so maybe they could lose weight, Lap-Band or some kind of procedure.

But in the meantime, maybe we'll recommend that you get these injections to see if you can get along until you can get a joint replacement. So there would be a question about efficacy in this type of a population, whether it's efficacious as it is in patients of this size, whether it would last as long, and I think that would be something that would reflect the type of use that would go on in the United States.

MR. HALPIN: I'd like to have Dr. Holmdahl come up and answer that question.

DR. HOLMDAHL: Thank you. I mean, that's a very, very appropriate question. So we have actually done that comparison ourselves. Slide on, please. This is the baseline characteristics of all patients

in the trial, in the column here, and we have compared that to the OA initiative cohort that is published with U.S. patients, and there is -- fairly consistent between the two cohorts, in terms of mean age and actually BMI, as well as, you can see from here, what the difference -- would rather be the ethnicity. That is what stands out. But the baseline characteristics, we otherwise believe, are very, very comparable.

DR. OLSEN: But these still reflect trials rather than clinics, and I'm just -- I mean, even that BMI of 30 is not -- you really do see BMIs of 44, and those are people who walk into the clinics, so I just think it's an issue that might be out there in clinical practice.

DR. HOLMDAHL: We have actually looked at the efficacy of the product in terms of BMI, whether patients have an increased BMI or have normal BMI, so I'd like to show that slide here. Slide on, please. And we defined that as increased BMI is greater than 25, and since normal BMI is -- 25 and we do not see any decrease in efficacy. What you see here is the WOMAC A score by BMI. We don't see a decreased efficacy in the patients with increased BMI.

DR. BLUMENSTEIN: What are the number of

patients on that -- in that split?

DR. HOLMDAHL: I would like to come back to you with that information. Does it say on the slide? I can't see that from here. It probably should say here on the slide, but I don't think the Panel members can read that either.

DR. EVANS: I think one thing you'd want to do is actually assess interaction of baseline BMI with treatment effect, with treatment. You probably don't have enough power to find anything, but that's really the way to look at it.

DR. MABREY: Dr. Olsen?

DR. OLSEN: Yeah.

DR. MABREY: I'll just add that I'm also from Dallas, as Dr. Olsen is, and as you all know, everything's bigger in Texas, including our clinics. Dr. Skinner.

DR. SKINNER: Thank you, Dr. Mabrey. My questions or concerns are similar to Dr. Goodman's and basically revolve around my concern that the two groups --

DR. MABREY: We're going to get a clarification of one thing first.

DR. SKINNER: Sure.

DR. MABREY: We're going to straighten it

up.

DR. SILLIMAN: Okay, great, thank you.

DR. MABREY: Sorry about that.

DR. SILLIMAN: Can I have that slide back again? Slide on. So we did actually look at some additional covariants that weren't pre-specified. BMI was one of the ones that we looked at, and there was no significant treatment by covariant interaction in this case. The p-value was .313.

MR. HALPIN: I'd like to have Dr. Polisson just comment briefly.

DR. EVANS: I think that's the way to look at it is through a de-interaction, although the power to find significant interaction is probably going to be pretty low, but I think it's what you can get out of it.

DR. POLISSON: So this is anecdote. What Dr. Olsen raises is a very good point about very large people, and we do have an investigator that works with us, Dr. Waddell in Louisiana, and he has used Synvisc in this very obese patients and claims that they, you know, have similar types of efficacy as you would see in people with a BMI that was listed in the results of our trial or the osteoarthritis initiative. We recognize that that's sort of a

shoot-from-the-hip anecdote, but that at least speaks from my experience and our experience with this particular physician.

DR. MABREY: Okay, Dr. Skinner, I'm sorry to interrupt you.

DR. SKINNER: Okay. As I was saying, my concerns are similar to Dr. Goodman's, and I think, although he didn't specifically say it, I think his concern is that the control group and the experimental group are the same group, statistically, anyway. And my concern comes in -- partially in making sure that while the WOMAC score and pain scores are significant in defining the group, their snapshot in time, and the Kellgren-Lawrence evaluation is more of a less time-dependent situation.

So I'd be interested in seeing that the Kellgren-Lawrence two-three group is similar in the two groups, the percentages for each one, because that would help reassure me that the two groups had the same amount of OA, and similarly to the comments Dr. Goodman had regarding varus/valgus.

Another issue is that the rescue medication was paracetamol or acetaminophen, and while that's an adequate rescue medication, I guess, the patients

were allowed to take nonsteroidal anti-inflammatory drugs with a half-life less than five hours. I don't know what that means. But I'd be interested, again, to know if the NSAID medication used by the two groups was similar because even though they wash out for 48 hours prior to their presentation for follow-up appointments, if they're not very active in that time, the effects of the NSAIDs can be carried over.

The third thing is that it's pretty well accepted, I think, by most orthopedic surgeons, although -- and rheumatologists, you'll find no one who says that they can't put the needle into the joint every time. This sometimes misses, as has been shown in the literature. Dr. Jackson in Long Beach did a big study on this and showed that he did much better in getting the needle into the joint if he used an image intensifier or X-rays.

And the criteria, I guess, for getting into the joint, in this study, was that there was aspiration, but I couldn't find data on how many of the knees actually achieved the successful aspiration. If it was similar in the two groups, again, I'd feel more secure that the two groups were similar. So I think those are my comments.

DR. MABREY: Does the Sponsor wish to

respond at this point or wait?

MR. HALPIN: Yeah, we would like to respond to the NSAID question.

DR. HOLMDAHL: So we did analyze the use of all concomitant medications -- slide on, please -- and which is shown here in this bar graph. We have the proportion of patients on the Y axis. This is all analgesics. This is where there was a slightly increased use in the control population.

Anti-inflammatory was the same proportion. There was also some patients who were taking aspirin. There was a slightly increased proportion in the control group. There were topical products for joint pain, which also were a little higher in the control population as was corticosteroids for systemic use.

This is also another, I think, interesting finding, where we had drugs for acid-related disorders by the two treatment groups and we had -- this difference is actually statistically significant, there is a greater proportion of patients in the control group taking drugs for acid-related disorders.

DR. SKINNER: To follow up on that, the protocol required them to take it only a certain amount of time per month.

DR. HOLMDAHL: Yes.

DR. SKINNER: This slide shows the number of people who took them or --

DR. HOLMDAHL: Yes.

DR. SKINNER: But it doesn't give an idea of how much they took?

DR. HOLMDAHL: That is correct.

DR. SKINNER: Do you have any information on --

DR. HOLMDAHL: I am hoping that I will be able to get back to you after lunch with more detailed information regarding that.

DR. MABREY: Thank you. Dr. Blumenstein.

MR. HALPIN: I'd like to have Dr. Simon come up and speak briefly about the KLG inclusion/exclusion criteria.

DR. MABREY: Okay.

DR. SIMON: I think it's important to remember that, in designing clinical trials for determining baseline characteristics for a pain trial as opposed to a functional outcome trial, one recognizes that you use K-L to define that a patient has established osteoarthritis as a disease state.

However, there's no evidence in any literature that correlates the extent of the pain

that a patient might have directly to what X-ray they actually have. How often do we clinically see people with eburnated joints who actually are able to do pretty well, and people with only mild disease, by X-ray characteristics, who are extremely uncomfortable and very complaining? So I believe it's really critical to understand the utility of K-L to define what in fact is going on with the patients. And as you can see on this slide -- slide up, please -- what was not shown on the demographic data previously in the core presentation is here is the K-L grade and you can see that, basically, there was an attempt to exclude people with Grade 4 and Grade 1, and basically it was a reasonable distribution of people with obviously established disease, which is really the only way that one can characterize the utility of K-L in the context of distinguishing patients from one group to another. Next slide, please.

And as you can see here, there is a slight difference between the ratio of K-L Grade 2 and 3 between the two groups. Neither interaction nor the K-L grade showed a statistically significant effect. Thus, in fact, the K-L grade was not helpful to understand how one group may have responded versus

another group, but at the same time allowed the appropriate patient to be recruited into the trial.

DR. SKINNER: And actually biased it against Synvisc?

DR. SIMON: To a degree, one might argue that that might be true.

MR. HALPIN: I think we'd like to answer the remaining questions regarding inclusion/exclusion criteria and aspiration of the knee after lunch.

DR. MABREY: Okay, thank you. Dr. Blumenstein, you had another question?

DR. BLUMENSTEIN: Yeah, a quick one. Could I see Slide CC-11? Could you describe the yellow study, please? Is it complete? What's its status?

MR. HALPIN: The yellow study is a completed evaluation of a different formulation of a viscosupplement, so it's not Synvisc; it's a bacterial HA-based viscosupplement.

DR. BLUMENSTEIN: Would that have the same indications we're talking about today?

MR. HALPIN: Yes, I believe it would have the same indication for use as what we're talking about today.

DR. BLUMENSTEIN: How large is the study?

MR. HALPIN: I think this study -- well,

the study design was somewhat different than the design we're talking about in that it was not a comparison to saline, so it was --

DR. BLUMENSTEIN: Is it a comparative study?

MR. HALPIN: It's a comparative study between the new viscosupplement and intra-articular steroid.

DR. MABREY: Anything else? I just have one thing I'd like the Sponsor to consider. You've reported that your study was -- your study subjects were 96 percent Caucasian, taken from sites throughout Europe. The U.S. Census Bureau 2006 data shows that only 66 percent of the U.S. population is Caucasian, with 15 percent being Hispanic or Latin American, 13 percent African-American, and four percent Asian.

In addition to that, there are recent studies and I'll reference one of them, Jing Song, et al., in *Arthritis Care & Research*, Volume 57, Number 6, August 15th, 2007, talking about the ratio among ethnic differences and activities of daily living, disability in older adults with arthritis, a longitudinal study, and in that they report that the ADL disabilities are likely to be twice that for

African-Americans and Hispanics than they are for Caucasians.

So when I look at these figures -- and again, I'm not a statistician. I did not stay at a Holiday Inn last night; I stayed here at the Hilton. I'm looking at one-third of the population, number one, has not had this study conducted, and yet they are twice as likely to be subjects of this device. And I'll let the Sponsor reserve that for that afternoon. I know it's a big chunk to bite off. And a second question --

MR. HALPIN: Thank you.

DR. MABREY: -- regarding the repeat trial at the end of the study, were there any efficacy results obtained from that, and why not?

MR. HALPIN: The repeat portion of the trial was a four-week study of the safety of repeat treatment, and the duration was not long enough to study efficacy endpoints.

DR. MABREY: All right, thank you. Well, it's exactly twelve o'clock. You've all done extremely well with keeping us on time, and for being so good, I'm going to call a one-hour lunch break and have us return here at one o'clock.

I will advise the Panel members, please,

you are not to discuss the subject matter at lunch. I'll also remind you that, for the Panel member, we have lunch in the restaurant, in a separate room, and that's meant to speed us through our dinner process. Please take any personal belongings with you.

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

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

(1:00 p.m.)

DR. MABREY: I'd like to call the Panel meeting back into session. If we could close the auditorium doors. We'll now resume our Panel discussion.

As a preliminary announcement, please, if you're speaking, please direct your voice to the microphone. Some of the people in the back are having a hard time understanding. I believe a lot of that has to do with the Sponsor because you're standing at a podium and you tend to back away. I don't know why you're backing away, but please get a little bit closer to the mike and the folks in back can hear you.

Is the Sponsor now prepared to answer the Panel questions from this morning?

MR. HALPIN: Yes, we are. I'd like to first have Dr. Lena Holmdahl come up and answer the remaining inclusion/exclusion criteria questions from this morning and also address verification of needle effusion and aspiration and speak briefly on the NSAID volume question.

DR. HOLMDAHL: I'd like to start with the last question first, regarding additional information

of concomitant medication, and the answer is that we didn't collect the data in such a way that it'll enable us to show, over time, if there were differences other than what I already showed you in the morning. We did collect data in the beginning, and then we only collected subjects on various -- the various concomitant medications. So I have no further information and that is the answer.

DR. SKINNER: Could I ask a quick question on that? Which NSAIDs were acceptable NSAIDs? Which ones have the half-life less than five hours?

DR. HOLMDAHL: Dr. Simon is going to address that.

DR. SIMON: As everyone here knows, there are many different nonselective nonsteroidals that are presently available in the United States as well as a COX-2 selective inhibitor. The ibuprofen is the one with the shortest half-life. It ranges between one and a half and two and a half hours, depending on the patient. So that is a serum half-life, not necessarily a biologic effectiveness half-life. And I presume that that's what you're referring to.

Now, I have no idea what that particularly meant for this particular trial, but that in fact fit that category. And as you know, that then reflects a

significant number of patients who take OTC nonsteroidals because that's one of those that is particular available. Obviously the other one that's available OTC in the United States, naproxen has a 13-hour half-life. Did that help?

DR. SKINNER: Yeah.

DR. HOLMDAHL: Then I'd like to go on to address verification of needle placement. And we asked the investigators to try to verify correct needle placement either by trying to aspirate joint fluid and to ensure that there were at least a couple of drops that they could aspirate. And if they couldn't do that, then they were asked to use their clinical judgment to ensure proper needle placement in the joint.

And the last question I would like to address are all the questions pertaining to inclusion/exclusion criteria and in particular to tense effusion and to deformities. The target here was to include patients with mild to moderate OA, which is the current indication for Synvisc, and as we have mentioned this morning, it has been on the market in the U.S. and worldwide for many years. So that was the target population.

So for that reason and also for the reason

that we were concerned that major deformities, in and of itself, could have an effect or ameliorate the effect of a viscosupplement, patients with major deformities were included. And the assessment of whether a deformity could have this impact or not was left to the clinical judgment of the investigator.

When it comes to the tense effusions, there was an exclusion criteria for tense warm joints, with a criteria of inflammation, and the reason for excluding those patients was that there is international recommendations and treatment guidelines that is recommending these patients to be treated with intra-articular steroids. So therefore we thought that that was appropriate to have that exclusion criteria and whether a joint was -- fulfilled these criteria or not was also left to the clinical judgment of the investigator.

MR. HALPIN: I'd now like to have Dr. Polisson and then Dr. Simon come up and speak to ethnic representation in the Synvisc-One clinical study.

DR. POLISSON: So the question was a great one, and I just like to start out by saying that -- to remind the Panel and everybody in the audience that this Synvisc-One that we're reviewing today is

not a new molecular entity. It's just a simple regimen change of putting Synvisc three-by-two into one syringe. So it's a product that's been out there for a long time, for 10 years. Four and a half million patients. We've got a lot of experience with it across races. And the Synvisc trials that have been done in the U.S., both as part of our initial application as well as other studies that have been done post-approval, really kind of reflect the same distribution that you're seeing here in the Synvisc-One program.

Now, let me have this slide on. So this is what we showed when we analyzed the response in WOMAC A across the time points by ethnicity, and these are, I acknowledge, incredibly small numbers but at least a trend in the right direction, and that is to say that the top two rows, if you look at the 10 non-Caucasians who are randomized six and four, you see a decrement in WOMAC A by a Likert scale of minus 1.54 in the Synvisc-One group and minus 1.01 in the control group. So to the extent that that says anything, I think, at least you know the data with respect to this particular trial.

Now, you raised a bigger question, though, and I think, as a rheumatologist, I don't know of any

biologic plausibility that there should be a difference in safety and efficacy with this type of therapy that -- and OA expression that would go across racial divides. That said, I would like to ask Lee Simon, who actually is more of an expert than I am on this particular area and has published in this area, to comment further, if I could. Thank you.

DR. SIMON: So it's very interesting to be able to address you about this particular issue. I'm an author on two of the most critical papers about the use of hyaluronic acid in the United States. One was in the *Rheumatic Disease Clinics of North America*, and the other one -- I was first author of that, and the other one is by Brandt, et al., in *Arthritis & Rheumatism*, both this decade.

And basically, we did extensive literature review on the entire field, and we attempted to understand the trial design issues that some of you have already brought up as it relates to what happens, one of which is continued rescue use throughout the entire trial, for example, how that can obfuscate benefit, and other issues that have plagued the particular field, one of which has to do with local therapy for two joints versus one joint.

But one of the things that really became clear in our study is that -- in our analyses of these data is that, A, there was no real differences in how people of different racial backgrounds responded to the kind of therapy. We would've pointed that out because we believe that that's an important issue. Two is, in thinking about a local therapy for pain, and being one of the people -- I was the author, one of the authors of the OMERACT-OARSI responder index, one of the problems in thinking about ADLs and responsiveness to therapy is whether or not a pain drug, an analgesic drug, can actually really alter function to the extent that you might want to see in a clinical trial outcome. And powering such a trial can be very difficult.

We're grateful that the FDA has actually chosen to ask sponsors to use the OMERACT-OARSI outcome responder index as secondary outcome so we can learn more about it, but we are a little bothered by how it's being interpreted.

So in the end, my comment really has to do with the fact that we really found, in an extensive review of the literature, any -- no real particular biases based on racial background, ethnicity background, in the context of outcomes in a highly

problematic field of outcome measurement with this kind of therapy. I don't know if that is totally helpful.

DR. MABREY: Yes, it is, thank you.

MR. HALPIN: And then I'd like to have Dr. Silliman and Dr. D'Agostino come up and address the model fit issue and multiplicity, briefly.

DR. SILLIMAN: Thank you. Let's see. So first I wanted to respond to -- I think it was Dr. Evans' question about the model fit for our primary model for the primary endpoint. So we fit that repeated measures, analysis of covariance, with an unstructured mean as well as an unstructured variance/covariance structure.

So in terms of assumptions, it was sort of the minimal amount of assumptions that we needed to make. We did check the residuals, and that plot of the residuals were fine. I can show that if you'd like. No? Okay.

And we also did some work when the FDA asked us to fit the model using site as a random effect. They also suggested that we pick a variance/covariance structure based on the AIC, the Akaike information criteria. So we did some work on that, which I can go through. Slide on.

So we fit the five different covariant structures here, the first auto-regressive moving average, first auto-regressive, spatial power, compound symmetry, and tuplets. And what you see here is that you have the AICs and you're looking for the smallest value. That's the covariant structure that gives you the best fit. You'll notice here that that was actually the first order auto-regressive moving average. Thank you. That's this one.

We were unable to get that model to converge consistently on the various populations, intent to treat versus protocol, as well as we were unable to get it to converge consistently on the secondary endpoints, so we therefore moved to the first order auto-regressive as our choice for the covariant structure, and that's what we used for all the FDA-requested analyses.

There was also a question about -- you can put the slide down, thank you. There was a question about whether we had done any nonparametric analysis. We actually did not do any nonparametric analysis for the primary endpoint here.

And then I wanted to maybe introduce Professor Ralph D'Agostino and the topic of multiplicity for the secondary endpoints. Slide on.

So this is the IMMPACT paper that Dr. Dworkin spoke about, and in there there's actually a statement about the lack of a need to adjust for multiplicity for secondary endpoints, and as Dr. Dworkin mentioned, this was an effort involving several FDA officials. Next slide. Slide on.

So this is just -- there's just two slides here with a quote from the paper, and then I'll introduce Dr. D'Agostino. So I bolded here that the statement from this paper was that, in a regulatory context, when there is a single pre-specified primary efficacy endpoint, and all additional endpoints are declared as providing only supportive or exploratory information, adjustment for multiplicity will typically not be necessary. And the reference here is actually the Committee for Proprietary Medicinal Products, points to consider on multiplicity issues in clinical trials document. This is part of the European regulatory authorities. Slide on.

And then it goes on to say, there are other circumstances in which multiplicity adjustment is usually not considered necessary, for example, to examing secondary hypotheses or secondary endpoints. And this is actually the reference to Dr. D'Agostino,

for his paper and stats and medicine on controlling alphas in a clinical trial, the case for secondary endpoints.

And as we heard before, Dr. D'Agostino is internationally recognized and very well published statistician, so I'd like to introduce Dr. D'Agostino.

DR. D'AGOSTINO: Thank you. The material that has just now been presented is pretty much what is to be said. Could you put the slide back on, please? The history of secondary endpoints is pretty long, and it does pay heed -- and we should have heed in terms of are we handling them correctly.

I mean, I've been around for a long while, as a number of other people here, and there was a time when you would run hypotheses tests for clinical trials and you would just give a long list of variables and whatever was significant you declared as your winner. Then there was a time when one would say, okay, let's separate primary from secondary, but it did make a bit of difference where the significance was where you declared winners.

And then there was a time when -- not that long ago, when things like mortality was being put in as a secondary variable in cardiovascular trials and

nothing else would be significant except the mortality, and that was being elevated to the claim. And the sorting out in the paper that is quoted there, the 2000 stat medicine paper was an attempt by myself and a number of other individuals, FDA included and a lot of FDA advisory committee members, to sort of sort out what the issues were.

And the bottom line is that a good trial should have a small number of primary efficacy variables, one if possible, and then some secondary variables. And if the secondary variables are -- well, first of all, the primary is where your money is, and if it's one variable that's in the primary, then you must show significance on that to go anywhere. If that is significant, then you can say, what about the secondary? And if the secondary, as it's quoted here -- and this actually -- could you go to the SM-3, please? This one here.

And actually the paper that is quoted from me has this also in it. If the point of the secondary variables is solely to give confirmation consistency to the primary, then there is no real need to control the alpha. Where you need to worry about controlling alpha in the secondary is if you have some secondary variable, again, after the

primary has been significant, is shown to be significant and there's some secondary variables that you would, say, for example, in a regulatory setting, you'd like to elevate to being part of the label claim and so forth.

And if you have that in mind, then it's very important to have the secondary variables a priori declared in that fashion, that you're going to look at them after the primary, you're going to look at them as possible variables to make claims with. And then you have to have very tight control of your alpha. We call it study-wide alpha. If, however, in our case, we're interested in these secondary variables as confirmation that, are they in the right direction?

And if you look, no matter what was done by the FDA, no matter what was done by us, the effect sizes, the direction, the differences, are all going in the same way. The drug is better than the placebo. And what we're trying to do with the study -- what the Sponsor is trying to do is say, here's the significance, no matter -- here's the primary. No matter how you look at the primary, there's significance. And do the secondary; go in the right direction.

And, in fact, they all go in the right direction, the ones on the WOMAC A1, where it's walking, pain on walking, the global variables. They all go in the right direction. Depending on which analysis you use, you get sometimes over .05, sometimes under .05, but they're all in the right direction, they're all in the same direction. That's the key to, I hope, the way you interpret the multiplicity. We're trying to show consistency.

As far as some of the procedures used, the agreement between the -- there is an agreement between the FDA and the Sponsor in terms of the primary. No one's questioning it. As a matter of fact, when the FDA looked at it, they even got a better level of significance. When you go to the secondary, there's a discussion about what's right and what's alternatives. CC-64. Do you have that one, by any chance? Can you pull that up?

If you look at this -- thank you. If you look at this here, the first column is what the Sponsor produced when they did the proportional odds model. Again, this was pre-specified, it was well thought out, and the analysis showed lots of consistency with the primary outcome.

The FDA, in looking at it, was trying to

make sure that there's a robustness to it, and what they basically did -- and going back to Dr. Evans' question there, they basically used like a nonparametric method. When I started out in statistics with the FDA, everybody was using what I call Likert scales. Everyone was using Likert scales and they were doing t-tests on analysis of covariance, and the question was were they really valid? And we have done a lot of work on it, showing they are in fact valid, they are robust procedures, they do give you appropriate alpha values. The problem is that, in terms of where the Sponsor is coming from, the WOMAC A, the PTGA, the COGA, these are variables that have small scales, and people have spent a lot of time asking about what's the better analysis. Can you do something better than just doing a t-test, just doing analysis of covariance?

And the proportional odds model came, and there were some very good questions about the assumptions. Our analysis, in terms of the assumptions being met, shows over and over again -- and again, as Professor Evans said, you're accepting a hypothesis, but there's no reason to believe the proportional odds assumption isn't met.

And we think that the first list of p-

values is the appropriate list, but even if you go to other procedures, look at that whole sheet there, everything is showing the same direction. Again, this is supported for consistency. Thank you.

MR. HALPIN: Those are all the responses the Sponsor has at this time.

DR. MABREY: Do the Panel members have any additional questions for the Sponsor or for the FDA?

(No response.)

DR. MABREY: Okay. At this time now, we can focus our discussion on the FDA questions. Copies of those questions are in the back of your FDA handout. For the Panel members, the questions that are in your three-ring binder have been changed a little bit, so go by the Panel questions that are in the slide handout.

Dr. Lee, would you like to read the first question for us?

DR. LEE: Yes. Chairman and Panel members, please note that Question 1 was modified to clarify the content of the previous Question 1.

Panel Question 1. Based on the mean difference observed between Synvisc-One and the phosphate-buffered saline control for the primary endpoint of the study as shown in Table 18 of FDA

Executive Summary, the group difference was 0.15 out of the five-point Likert scale. Please discuss the clinical relevance of the 0.15 incremental advantages of Synvisc-One over the control in the mean difference in change from the baseline for the proposed indication for use.

DR. MABREY: Dr. Evans?

DR. EVANS: I guess I sort of have mixed feelings about the clinical relevance of -- that is seen. I thought Dr. Dworkin's presentation actually shed some light on it. I think, from a statistical standpoint, I was actually encouraged by the consistency of at least the sort of statistical significance and the similarity of effect sizes in various analyses.

So I think, from -- you know, as you evaluate treatment effects in clinical trials and you're looking at statistical significance and you're looking for clinical relevance, I felt -- I sort of feel somewhat encouraged by -- from the statistical standpoint of the statistical significance in sort of consistency of effect sizes in the sensitivity analyses across models that were fit. I have a little bit more trouble trying to interpret the clinical relevance of the effect size. I think it's

sort of a clinical question.

I thought Dr. Dworkin, you know, he actually had a list of considerations to look at when trying to make a decision about what would be clinical relevant, and I think there was a couple of issues there. There was also -- he also alluded to a document that basically said that any effect of -- between group difference effect would be relevant in some way and not to try to -- and he made clarification not to confuse group differences with what would be relevant for -- relevant changes for individual patients. So I think that the clinical relevance question and part of Dr. Dworkin's list was to sort of consider it, to look at the effect sizes and interpret them within the context of secondary variables, within the context of the safety data and what it sort of costs and risks to, you know, what the other costs and risks and benefits are associated with the therapy.

But I'm still a little unclear about how to interpret the clinical relevance. I feel a little bit more confident about -- or a little more encouraged about -- from a statistical standpoint.

DR. MABREY: Dr. Goodman?

DR. GOODMAN: I think that this question

and Question 3 are basically the crux of the decision-making process. I admit that when I first went through this manual, I thought a difference of .15 was really negligible on a five-point scale. However, I think that we've been presented with comparable data from other interventions which shows that that is the same level, approximately, of other interventions that we use in the clinic. So I was encouraged by that.

I also was very happy that I do total joint replacement because I think that's probably the most effective of any intervention that there is. And that's all I have to say. Thank you.

DR. MABREY: Thank you. Dr. Olsen?

DR. OLSEN: In terms of the clinical relevance, maybe I'm a little more able to judge that from the statistical aspects of this, but I think my context is that -- sort of like what was just brought up here, that these patients are looking at a longstanding problem with minimal significant alternatives.

There's very good safety profile to what they got. They didn't drop out, but of course the saline-injected people didn't drop out either, but I think it's kind of a measure of they were all hoping

that there would be something here that would help them.

So I put it in that context, and it would be something that we would say to a patient, there is a -- that it would have to be described this way, but it is not something that is going to change your life overnight, but it might extend you to the total joint replacement or have some other limited goals. And in that sense, I think it does have clinical relevance to have even a small degree of improvement.

And I was encouraged by the fact that all the other markers seemed to go in the same direction, so there wasn't anything else there that seemed to suggest some underlying current moving in an opposite direction. They were all going in the same direction.

I'm not concerned about the difference with the projected number versus the number that came out because I think that's all based on assumptions that aren't always -- I mean, it's interesting to me that you assume that 25 percent of people would drop out and so few people dropped out. So you know, your assumptions, you try to be real careful about them, but that's one that you didn't have to be that careful about.

So I don't know what we learned from that, but there's something kind of interesting about this is what happens when you do these things to patients. So the bottom line is I think it's small but probably clinically relevant that this would offer something in a field where there's limited choices, so I don't have a lot of concerns.

DR. MABREY: Thank you. Dr. Skinner?

DR. SKINNER: Well, my comments aren't a lot different from the other two Panel members. I think everybody in the room acknowledges that the improvement with this injection process is modest, and this is one of those things that'd be kind of no-brainer if it was a $50 injection. But when you add an order of magnitude to that, it makes the clinical relevance more significant. It'd be nice if it was a nice, inexpensive drug that you could give once and get this much clinical improvement.

Based on that, I think that there is clinical improvement, and the modest effect is clinically relevant. It's just a shame that it's so expensive.

DR. MABREY: I understand that we're not to consider the cost in --

DR. SKINNER: Of course not.

DR. MABREY: -- our deliberations, but I appreciate your comments on that.

DR. SKINNER: I didn't consider it at all.

DR. MABREY: Dr. Blumenstein?

DR. BLUMENSTEIN: I have a wish list, actually. To me, the .15 isn't so relevant as the .97 that was used to plan the trial. In other words, it was -- somebody declared that the alternative, the specific alternative hypothesis to be used to compute the trial size would be based on a difference of .97 -- 297, .297. And I assume that number is comparable to the .15, if I'm understanding all of the numbers that are being thrown around here.

So the company then did a trial and they gathered data, and the data has said that you should reject the null hypothesis in favor of the alternative hypothesis, and it was planned with that .297. So to me it's the .297 that has more meaning than the .15. And I have no basis for understanding what that number means. It seems small to me, but I'm listening keenly to my clinical colleague.

The other piece of wish list -- on my list wish is that I sure do wish I had some kind of a comparison either between placebo and the three treatment, or between the one treatment and the three

treatment because I feel like I don't know where I am. And that's all I have to say.

DR. MABREY: Thank you. Ms. Rue?

MS. RUE: In the clinical relevance, I feel, from the discussions that we've had on how effective it was, is significant. But also, I think the clinical relevance, as far as a consumer basis, is how it changes access and availability with the only one injection instead of three and how this impacts their life, as far as their work-related and other things that they have to change, and it's different only having to do it once as opposed to three times.

DR. MABREY: Thank you. Comments?

MS. GEORGE: Well, obviously, I have the least clinical expertise here, so the clinical aspects aren't really key for me. But a couple things that came to mind when I would listen to everybody talk about this was, number one, I believe I remember seeing that it was actually the FDA that wanted the comparison to a placebo, not to the existing, so that's one of the reasons why that data is not available to us in this. I think if we had wanted to look at how the original was, that would've been available in the other original PMA, I would've

expected.

But the other thing that came to mind for me was that whole aspect of going for three shots versus one. Just as a patient, I would think that patients would be much more apt to show up that one time and -- rather than, you know, three times and the time, and I think one of the speakers this morning actually brought that up as well. So that's all I have to say.

DR. MABREY: Thank you. Mr. Melkerson, with regards to Question 1, the Panel generally believes that the statistics appear to be appropriate and well handled and that the clinical relevance of .15 seems to be acceptable, although small. Contrary to that, the Panel also has some concerns about the clinical relevance of this difference and also concerns about the selection of the cutoff of null hypothesis. Is this adequate for the FDA?

MR. MELKERSON: Yes, it is, thank you.

DR. MABREY: Thank you.

DR. LEE: Panel Question 2. Multiple secondary endpoints were tested without adjusting for multiple comparisons. Please comment on the adequacy of the applicant's analyses for the secondary endpoints in light of there being no pre-specified

multiplicity adjustment to control the overall Type I error rate.

DR. MABREY: I think I'm going to start with Dr. Blumenstein on this one.

DR. BLUMENSTEIN: Well, I accept the notion that there was not an intent to put forth these secondary endpoints as claimed to be included in the label. Nonetheless, I feel that I can't look at that collection of secondary endpoints without making some adjustment in my own mind, and I think that everybody else knows enough to do that, at least I hope they do, especially since these are highly.

In other words, you shouldn't be counting the numbers of significant secondary analyses that are significant according to a .05 criterion. That would be an incorrect way of assessing those secondary endpoints. I think the direction, as has been pointed out, is the most important thing. I'm still a little mystified by the fact that the SAP made such a clear statement about the importance of one of those secondary endpoints, and it has not been consistently represented or carried forward, and I can't help but wonder if the lack of significance of that endpoint isn't the reason that it's not being carried forward. So we have a lot of post hoc

analysis going on here.

DR. MABREY: Ms. Rue?

MS. RUE: I don't have any comment.

DR. MABREY: Ms. George?

MS. GEORGE: I think the only comment I have is, is again, we should remember what Dr. D'Agostino stated about the secondary endpoints, is that they are there as a support if the primary endpoint is met, and we wouldn't be here if the primary endpoint hadn't been met.

DR. MABREY: Thank you. Dr. Evans?

DR. EVANS: I guess, as a statistician, we always worry about multiplicity issues and multiple testing, but I think the key is how those tests are used and then, essentially, that they're interpreted correctly. I agree with -- I actually agree with Dr. D'Agostino. I think that the way I looked at and reviewed the results of this trial was to view the secondary endpoints and interpret those as -- to help assess the consistency of the effect and put the sort of effects of the primary endpoints into perspective, and that the claims are not necessarily being made on secondary endpoints, and in general, that's sort of the way I viewed them. I think the key is how you interpret things. Whether I make an adjustment to --

if I do a hypothesis test and I get a p-value and I make an adjustment to that because maybe I did two tests instead of one, well, the level of evidence is the same.

I'm just changing the bar on how I interpret it. I mean, the data has changed at all. So it's all about how it's interpreted and you realize that the more tests you do, the more chances you are of perhaps finding a false positive error. At the same time, I don't think there's a need to control alpha, necessarily, for every test. I think, as long as you realize that you've done a number of tests, there's a chance of potentially making a false positive claim, but you realize that that's important.

I think there needs to be thought about when do you need to control error for each test versus when you can sort of just realize that you've made multiple tests and make that adjustment. In this particular case, I'm not sure there's a need to control for the multiplicity involved with the secondary endpoints. In addition, I think that even if you make an adjustment for the secondary endpoints for the number of secondary endpoints that are being made, the adjustment's going to be fairly small and

in the sense that I don't think, qualitatively, I would change the way in interpreting the data based on the adjustment I would make for multiplicity.

And let me just make this point because I think this is perhaps one of the biggest confusions or misinterpretations of statistical output that is made in the literature today, is that there's an over-interpretation of p-values when we get -- and what I mean by that is both when a p-value is significant and when it's not significant. A p-value is a composite statistic. It's partly effect size, it's partly sample size, it's partly variation. And if you get a high p-value or you get a low p-value, you've got to find out what's driving it. It could be any one of those three factors that's driving it.

So I think oftentimes we spend too much time. You know, I think, as evaluators, sometimes we spend too much time worrying about whether we get under this magical 05 level. And I think people who are doing research spend too much time worrying about how to get under that 05 level and don't worry about trying to interpret what the data are telling you, and I worry about sort of that over-interpretation of p-values. And the only way to deal with that is to look at effect sizes through use of confidence

intervals, to perform sensitivity analyses through varying assumptions and missing data and things like that. I think, in this case, I have less concern about the multiplicity issue with the secondary endpoints because I do view them as sort of -- to look at them as consistency of effect and to help put the overall effect of the -- you know, of the intervention into perspective.

DR. MABREY: Thank you. Dr. Goodman?

DR. GOODMAN: I've nothing further to add.

DR. MABREY: Dr. Olsen?

DR. OLSEN: I didn't have any concerns.

DR. MABREY: Dr. Skinner?

DR. SKINNER: Nothing further to add.

DR. MABREY: Mr. Melkerson, with regards to Question 2, the Panel generally believes that secondary endpoint analysis was appropriate and that there is probably no need to control for the secondary endpoints, and that even if adjustments were made, they would be small, anyway. Having said that, the Panel also has -- does have some concerns about use of secondary endpoints and post hoc analysis. Is that adequate for the FDA?

MR. MELKERSON: Just a point of clarification, and maybe it's aimed at Dr. Evans and

Dr. Blumenstein. In terms of interpretation or limitations or qualifications and of presenting the secondary endpoints, any suggestions on how you would present that type of information, given the concerns of multiplicity?

DR. BLUMENSTEIN: I'll respond to that. I think it's really quite simple, that they have told you that they're not making any claims, so it doesn't need to be in the label. So you'll have a short label to write here, if there's final approval.

DR. EVANS: Yeah, I think if the question is directed at labeling, that's probably the right approach. I think, in terms of if a report is generated and as we try to make others better understand the data, that there's one clarity of how many tests were performed, that this is perhaps a statement about something to the effect of, even if this intervention has no effect whatsoever, I would expect to see so many of these tests, X number of these tests potentially show false positive results. And that's just an expectation, but it helps put into perspective, you know, what you would expect to see. And so I think part of the multiplicity problem is just clarity of reporting about how many tests are you looking at, what significance level are we using,

how many would I expect to see significant even if there was nothing going on, even if this was just no better than placebo, and that there's clarity of that.

But I think, in terms of labeling, I think, in consistency with what I said earlier, the reason that I'm not worrying about the multiplicity issue, to be consistent with that, I think the answer, as Dr. Blumenstein said, is that it doesn't go into the label because you're not controlling for that specific effect.

DR. MABREY: Thank you.

DR. LEE: Panel Question Number 3. Under 21 C.F.R. 860.7(e)(1), effectiveness is defined as reasonable assurance that, in a significant portion of the population, the use of the device for its intended uses and conditions of use, when accompanied by adequate directions for use and warnings against unsafe use, will provide clinically significant results.

Considering the study design and endpoints discussed today, please discuss whether the clinical data in PMA/Supplement provide reasonable assurance that the device is effective.

DR. MABREY: Dr. Goodman?

DR. GOODMAN: Well, I was impressed with how clinically effective or ineffective most of our conservative treatments are for osteoarthritis, and this device is no more effective than some of the other alternatives. It is statistically more effective than aspirating and a placebo injection. So I think it is modestly effective, and that's about all I can say.

DR. MABREY: Dr. Olsen?

DR. OLSEN: Well, I believe the data show that it is effective. My hedge on this one is the definition of the population because I still have some concerns that the population in this protocol had a lot of differences with the population that people like us treat in this country, in terms of race and ethnicity and social status and size, body size, that I brought up before.

So I think, given the small numbers we're talking about here and how effective this was, those are variables that maybe the statisticians will agree with me, if you put other variables into the population, maybe we'd get a different outcome. So I have some concern about that. It's a minor concern because I still think it's being shown that it is effective, but that's my asterisk on that.

DR. MABREY: Dr. Skinner?

DR. SKINNER: I basically agree with Dr. Goodman. I think that it shows modest -- there's modest effectiveness.

DR. MABREY: Thank you. Dr. Blumenstein?

DR. BLUMENSTEIN: I agree that the statistical criterion has been met. I was particularly comforted by what I consider the correct model, where clinical site is a random effect, going in the direction that it did. So I think that we have met the statistical criterion on the study.

DR. MABREY: Ms. Rue?

MS. RUE: I don't have anything else to add.

DR. MABREY: Okay. Ms. George? Mr. Melkerson, in regards to Question 3 --

DR. EVANS: I think the effect is nonzero. Whether it's clinically relevant, as we've discussed, is a more difficult issue.

DR. MABREY: I'm glad we waited for your comment. Mr. Melkerson, in regards to Question 3 now, the Panel generally believes that the device is modestly effective, at least nonzero, but they do have some concerns about the nature of the population for which the device would be applied to. Is that

adequate for the FDA?

MR. MELKERSON: Yes, it is, thank you.

DR. MABREY: Thank you.

DR. LEE: Panel Question Number 4. Under 21 C.F.R. 860.7(d)(1), safety is defined as reasonable assurance, based on valid scientific evidence, that the probable benefits to health under conditions of the intended use, when accompanied by adequate directions for use and warnings against unsafe use, outweigh any probably risks.

Considering the adverse events for the device, please discuss whether the clinical data in the PMA/Supplement provide reasonable assurance that the device is safe.

DR. MABREY: Dr. Olsen?

DR. OLSEN: I think the data support, with reasonable assurance, that it is safe.

DR. MABREY: Thank you. Dr. Skinner?

DR. SKINNER: I think it's also safe. I agree with Dr. Olsen.

DR. MABREY: Dr. Blumenstein?

DR. BLUMENSTEIN: I concur.

DR. MABREY: Ms. Rue?

MS. RUE: I concur.

DR. MABREY: Ms. George?

MS. GEORGE: I concur, especially since the material has been out there for 10 years in the U.S. and 16 years worldwide.

DR. MABREY: Dr. Evans?

DR. EVANS: I agree.

DR. MABREY: And Dr. Goodman?

DR. GOODMAN: I concur.

DR. MABREY: That was easy. Mr. Melkerson, in regards to Question 4, the Panel believes unanimously that the device is safe.

MR. MELKERSON: Thank you.

DR. LEE: Reminder. The discussion of a post-approval study prior to --

MR. MELKERSON: Kevin, this question only comes up if there's a question regarding the Panel. Is that correct? I'm looking around.

DR. JEAN: I believe we can generally discuss this, hypothetically, at this point.

MR. MELKERSON: Okay.

DR. LEE: Reminder. The discussion of a post-approval study prior to a formal recommendation on the 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 premarket data submitted to the Agency and discussed today must stand on its own in demonstrating a reasonable assurance of safety and effectiveness in order for device to be found approvable.

PAS Panel Question Number 5. The applicant did not provide a post-approval study plan in the original PMA/Supplement. However,

(1) the clinical study supporting this PMA/Supplement was conducted in Europe and patient's characteristics may be associated with the treatment effects of the device.

(2) The follow-up of this PMA study was 26 weeks for the initial phase and 4 additional weeks for the repeat phase, while intra-articular injection of similar devices has demonstrated the treatment effects extended to 12 months after the injection.

(3) The literature has suggested that cross-linked hylan G-F 20 used by Synvisc may be associated with increased risk of severe acute inflammatory reaction. The exact mechanism of this association and its long-term consequences remain unclear.