presentation, at a dose of 200 mg per m^2, amifostine was well tolerated. There were no new or cumulative toxicities that were identified during the course of this trial. And nausea, vomiting and hypotension do remain as the most common adverse side effects attributable to the drug.

To wrap this all up, we found in the WR-38 trial that amifostine significantly reduced the incidence of greater than or equal to acute xerostomia and late, long-term xerostomia. Saliva flow was preserved to a significantly larger extent in those patients who received amifostine with their radiation therapy. Patient assessment via the patient benefit questionnaire also indicated clinical benefit to the patients. Amifostine did not reduce the anti-tumor efficacy, irrespective of whether we looked at local regional control, disease-free survival, or overall survival, and amifostine is safe at the recommended dose.

I would now like to turn the podium over to Dr. Gary Koch from the University of North Carolina at Chapel Hill, who will provide additional analysis

and insight into the statistical aspects of this trial.

DR. KOCH: Thank you very much. Not very often you have faculty from Duke and North Carolina working together. But it can happen and often does happen and happens quite successfully. I have assisted U.S. Bioscience as an independent statistical reviewer for the findings of WR-38. Since the statistical results for xerostomia, as presented by Dr. Brizel, were clearly convincing, I have given primary attention to anti-tumor outcome.

Showing non-inferiority of amifostine for anti-tumor efficacy was a primary objective of WR-38. And accordingly, it had well-planned analyses. I requested some additional assessments to confirm the robustness of findings for non-inferiority from the planned analyses. These assessments were based on the Kaplan Meier survival curves which were included in the WR-38 study report, the briefing book, and Dr. Brizel's presentation. They are not new, but they enhance understanding for what you have already seen.

Their purpose was to clarify how the

overlapping or better nature of the Kaplan Meier curves for amifostine than for RT alone convincingly supported the non-inferiority of amifostine for anti-tumor outcome. Let us now briefly review the statistical results for anti-tumor outcome.

The analysis plan for WR-38 prespecified the ratio of local regional control proportions at 12 months as the primary criterion. And these are the proportions that are shown here. The lower one-sided 95 percent confidence limit for this criterion exceeded the prespecified threshold in the protocol of .70 for non-inferiority, and that is demonstrated here and here for a one-sided confidence interval and here and here for a two-sided confidence interval. Moreover, these lower limits exceeded .80 as opposed to the specified criterion of .70. And we can see this for the rates at 18 months as well as those at 12 months.

Now these analyses at 12 or 18 months included 127 or 126 of the amifostine patients, and they included 135 or 133 of the control patients. Thus, losses to follow-up or censoring are relatively

small. That is, there are about 41 or 44 patients or 15 percent of the original 303 patients in the combined groups, rather than being high as possibly suggested in the FDA review. There were not that many patients lost to follow-up during the 18-month follow-up period shown here. So censoring tends to be low as censoring is usually understood to be. Also, over 50 percent of the patients, 77 in the amifostine group and 85 patients in the RT alone group, completed 18 months of follow-up with LRC maintained, and they should not be regarded as having censored time to event data, since they had the most favorable outcome of maintaining LRC for this entire follow-up period. Also, my understanding is that LRC rates tend to decrease slowly after 12 to 18 months, that is, most failures occur before 18 months. And so 12 to 18 months is a reasonably mature follow-up period for its evaluation. As stated previously, censoring for this mature follow-up period was relatively low, less than 15 percent. And so lack of maturity for follow-up data and excessive censoring are not limitations of this information.

Now a very important point in the FDA review was that evaluation of anti-tumor efficacy should account for all time to event data for all patients. And this is best done with the Kaplan Meier curves for the two groups, and they can be seen on the next display as clearly overlapping. And so these overlapping Kaplan Meier curves provide very clear evidence of what the similarity of maintaining local regional control is for the two groups. Now the FDA review has expressed some concern that the hazard ratio, as shown here, has a lower confidence limit of .64. But a hazard ratio is a very difficult concept to interpret in terms of similarity of survival curves like Kaplan Meier curves. The hazard function can be thought of as the negative slope of the logarithm of the survival curve. So a hazard function is a negative of the slope of the log of a survivorship function. That is a hard concept to understand, and the hazard ratio is actually the ratio of two such things. It is much more informative to understand non-inferiority in terms of what the Kaplan Meier estimates are doing. So we can look at the ratio of

Kaplan Meier estimates at 12 months or 18 months, and note that their lower confidence limit exceeds .89 here or .81 here. And this is a much clearer way to understand the non-inferiority or the similarity of patterns of survival over the full time course. It is also compatible with what the confidence interval on the hazard ratio indicated. And if you are interested in further understanding a hazard ratio, I have a display on a transparency that I could give comments about on that later. But I would rather you focus on the similarity of the Kaplan Meier curves, as I will proceed in the remainder of this discussion.

So in the next display, we have the Kaplan Meier curves for disease-free survival. Again, we note that their lower confidence limits are again above .90 over here for the ratio of the 12-month values, above .80 for the ratio of the 18-month values, and then on the next display for overall survival, we see the overall survival is clearly better for the amifostine group. And again, the lower confidence limits are in the vicinity of .99 at 12 months, .98 at 18 months. Very convincing information

for non-inferiority of amifostine relative to the control group.

And now if we proceed to the next display, ratios of even Kaplan Meier rates are a hard concept to understand. A usual way to proceed is to focus on the difference in success rates. So what we proceeded to do was to produce confidence intervals for the difference in the control rates as originally obtained in the analysis plan, and their lower confidence limit is -10 percent here and in the vicinity of -15 percent. And so we know that the difference in control is no greater than 10 percent based on 12 months and no greater than 15 percent based on 18 months.

Now we can help to understand these quantities by recalling guidelines that apply to anti-infective drugs. For anti-infective drugs, non-inferiority is established if the difference in success rates does not exceed 10 percent when the better of the two treatments has a 90 percent success rate. Here we have the 10 percent difference achieved by rates that are in the vicinity of 71 percent.

Also, if it turns out that the better of the two treatments has a rate between 80 and 90 percent, then the confidence bound decreases to -15 percent, which is satisfied here by rates in the 60 percent range. And if the better of the two treatments has a success rate lower than 80 percent, then the guideline for anti-infectives allows a non-inferiority margin of 20 percent. And so what you see here is these rates which are below 80 percent actually satisfy criteria that are substantially stronger for non-inferiority than the guidelines that are usually emphasized for anti-infective drugs.

Now if we proceed to the next display, again we see for LRC rates from the Kaplan Meier curves a lower confidence limit of -10 percent and -15 percent, and on the next display for disease-free survival, a lower bound of -6 percent or -10 percent. These are very convincing for non-inferiority. And on the next display for overall survival, we see lower bounds of -1 percent. We almost start to see a pattern where there is almost superiority in overall survival. Very clearly there is non-inferiority for

overall survival. And so in the final display, we note that lower limits of one and two-sided, 95 percent confidence intervals are indeed sufficiently high to clearly assure non-inferiority of the amifostine group for anti-tumor efficacy. Thank you. I would now like to introduce Dr. Russell, Senior Director of Clinical Research at U.S. Bioscience.

DR. RUSSELL: I would briefly like to present some data which supports the ability of amifostine to reduce the radiation-induced xerostomia and in addition summarize the results of two Phase III studies which support the conclusion that amifostine does not compromise the anti-tumor efficacy of the co-administered cytotoxic therapy.

As you can see on this slide here, this slide illustrates the number of studies that have been undertaken with amifostine where the incidence of xerostomia or reduction in salivary gland function has been looked at as an endpoint. These studies have been undertaken in an total of 554 patients. Dr. Brizel has outlined the details of study WR-38 in some detail, and for the purposes of time, I will only

concentrate on the study undertaken by Dr. Antonadou.

This slide outlines the treatment schedule of the study that she undertook. All patients received radiation at standard fractions of 2 Gray a day, to a total dose of 60 to 74 Gray. Patients received weekly carboplatin at a dose of 90 mg per m^2 per week. Patients were randomized to receive this regimen alone or this regimen preceded by a daily infusion of amifostine administered at a dose of 300 mg per m^2 prior to each fraction of radiation or the carboplatin respectively.

The patient demographics are outlined in your briefing book. But suffice to say that there was no statistical differences in the treatment arms for age, gender, tumor site, tumor state, or nodal status.

Here are the results of late xerostomia documented at 3 months following treatment. As you can see, those patients who received amifostine, there was a significant reduction in the incidence of Grade 2 or higher late xerostomia, falling from 83 percent in the control arm to 27 percent in the amifostine arm. And this result is highly significant.

We have recently received some updated information from Dr. Antonadou at 9 months and at 12 months for this endpoint. As you can see, the results that she reported at three months are preserved at these later time points, with reductions from 72 percent to 17 percent at 9 months and 56 percent to 8.3 percent at 12 months, both of these results attaining statistical significance.

She also assessed anti-tumor efficacy. And as you can see, using this regimen, all patients sustained a response to treatment. There were 20 out of 22 complete responses in those patients who received amifostine and 18 out of 23 patients in those patients who received radio-chemo alone. Local regional control at 18 months has been preserved with local control of 83 percent in the patients who received amifostine and 76 percent in the patients who were in the control arm.

The safety outline in the study showed that one percent of patients who received amifostine suffered from some nausea and vomiting, and all patients in this study received a 5HT3 antagonist

prior to treatment. Transient hypotension was recorded in 3 percent of patients, but in no patient did they suffer any long-term sequela.

So in conclusion, for this study there was significant reduction in Grade 2 or higher late xerostomia, preservation of anti-tumor efficacy, and amifostine was well-tolerated. And so this small study does support the findings that Dr. Brizel outlined earlier for study WR-38.

I would now like to move on to the supporter studies for anti-tumor activity. The first study was undertaken in patients with rectal cancer by Liu, et al. All patients in this study received whole pelvic radiation plus or minus amifostine at a dose of 340 mg per m^2 administered prior to each fraction of radiation, which was 2.25 Gray per day. All of these patients had metastatic unresectable cancer.

As you can see here, there is complete response rate to treatment, with 16 percent in those patients who received amifostine and 10 percent of those patients who received radiation alone. Median survival for this patient population was 15 months for

those patients who received amifostine and 12.6 months for those patients who received radiation alone, and these results are represented on the Kaplan Meier curve here.

The last study I would like to present is the ovarian cancer study that formed the basis of the previous approval. In this study, all patients received cisplatin and cyclophosphamide plus or minus amifostine administered at a high dose of 9 to 10 mg per m^2 prior to each dose of cisplatin. As you can see, with a median follow-up of 41 months, the survival curves are completely superimposable.

So we have demonstrated in three randomized, well-controlled, Phase III studies that amifostine does not compromise the anti-tumor efficacy of the co-administered treatment. I would now like to hand over to Dr. Walter Curran, whose is Professor and Chairman of Radiation Oncology at the Kimmel Cancer Center, and in addition Chairman of the Radiation Therapy Oncology Group.

DR. CURRAN: Members of the committee and the FDA, I just want to make a few brief comments

related to my observations regarding the ability of this agent, amifostine, to protect against the subsequent development of radiation-related xerostomia in head and neck cancer patients.

If you can have the slide off, I just have a couple of slides I will present later in my comments. I am a practicing radiation oncologist whose practice includes head and neck cancer patients, and I also serve, as Dr. Russell mentioned, as the Group Chairman of the RTOG. This is a cooperative group that seeks to improve outcome among patients afflicted with six major diseases among adult cancers, and some of our studies include ways to reduce or modify treatment-related toxicity.

I just want to say a couple of additional comments regarding xerostomia. The minimum radiation dose used to treat head and neck cancers is 50 Gray. And if salivary gland is included within the radiation field, that dose is largely lethal to salivary gland function. Since the majority of head and neck cancer patients receive radiotherapy as part of their management, most head and neck cancer patients are at

risk for xerostomia.

Xerostomia, as you have already heard, can be both permanent as well as devastating, causing problems with fissures in the lips and mucosa, difficulty with swallowing, the need for oral comfort agents, and really a need to change the lifestyle of patients afflicted with this. There is no approved therapy to reduce the risk of radiation-related xerostomia. There is one approved therapy for the management of xerostomia once it is established. However, that therapy requires frequent dosing and in most cases patients still need to take large quantities of liquids without really a satisfying effect.

It is clear that all of us who treat such patients need some other effective approach to reduce xerostomia. Now if you look at WR-38 as presented by Dr. Brizel, it is my view that it presents convincing data that Grade 2 or worse xerostomia is reduced by the application of this agent during radiotherapy. Many of us who followed this study and did not participate in it had hoped for such an effect, but I

was personally surprised at the magnitude of effect of the agent in this study. Many of you are familiar with the difficulty in reducing or modifying cancer therapy-related toxicity, and it was certainly gratifying to see such a positive result in this study. And this data, along with the data presented by Dr. Russell, certainly provides a strong core of evidence to support this agent's use in reducing xerostomia risk.

I want to spend a couple of minutes just discussing the issue of local/regional control, which was presented by Dr. Koch and Dr. Brizel as well. Clearly we see that there was an equivalent local/regional tumor control rate between the amifostine-containing arm and the control arm. But to further establish the fact that anti-tumor effect was not compromised by amifostine -- if I could have the first slide, please. Tom Pajak, the chief biostatistician at the RTOG did an analysis of over 500 patients who were entered into RTOG trials in the past. These patients had similar pre-treatment characteristics as patients entered into WR-38. And

what is shown on this figure is the local regional tumor control rates with the yellow figure showing the RTOG data base with the one-year local/regional control rate of 68 percent, 62 percent at two years. The yellow numbers down here are the numbers from the WR-38 study, in which one-year and two-year local regional control rates were nearly identical at 71 and 61 percent respectively.

Now the other issue that I think is relevant is whether adequate follow-up of the WR-38 patients has been done in order to assure ourselves that we have seen the patients that are at risk for local/regional failure. This is from the same RTOG data base of over 500 patients analyzed by Dr. Pajak. And it looks at at what time point patients suffered local/regional tumor failure. Out of all those patients who suffered a failure, 78 percent in this cohort of over 500 patients suffered it during the first year and over 90 percent during the first two years. And when you consider that in WR-38 the median follow-up time available is 26 months, the minimum follow-up time 18 months, clearly we have those

patients at an adequate time point to see what their risk for failure was.

So in summary, it is my view and the colleagues involved with head and neck radiotherapy that I have spoken to that the data regarding protection from xerostomia is convincing from WR-38. It is also a -- there is also a safety profile of the amifostine, which appears manageable from our point of view. Keep in mind that the doses of amifostine necessary to reduce the risk of xerostomia are less than the doses required to reduce the risk of cisplatin-related nephrotoxicity. And the RTOG has found the profile emerging on amifostine sufficiently interesting that it is currently testing this agent to reduce toxicity in two other disease sites.

DR. OSTER: I would now like to conclude our presentations, and I promise I will be very brief. The supplemental NDA for amifostine which we presented to you today represents a new indication for a drug which is already approved in cancer. Amifostine gives the opportunity to reduce the incidence of a severe, irreversible morbidity, xerostomia. Our pivotal

trial, WR-38, plus the supportive body of evidence contained in this SNDA confirms the efficacy of amifostine against the toxicity of xerostomia. All studies consistently report positive xerostomia results with amifostine.

In study WR-38, we showed with several different endpoints statistically highly significant results for acute xerostomia, late xerostomia, saliva production, and the patient benefit questionnaire instrument. These independent endpoints showed a strong correlation amongst each other. We believe that these findings are clinically meaningful and provide the patients with a true clinical benefit.

The studies contained in this SNDA also provide reassuring evidence of safety in our radiation program. Indeed, the toxicities with amifostine we observed at the dose used in radiation were substantially less in incidence and severity than for the higher doses typically used in our labeled indication in chemotherapy.

Two randomized studies with 100 patients and 242 patients were available for the initial

approval showing no evidence of tumor protection. The relevant follow-up endpoints chosen in these two studies were survival. WR-38 is the third well-controlled clinical trial, comprising more than 300 patients. Tumor outcome measures included local/regional control rates, local/regional control over time, disease-free survival and overall survival, all of which show no evidence, not even a hint, of reduced anti-tumor activity. We believe the clinical data presented today are reassuring that amifostine indeed preserves anti-tumor activity.

Amifostine was designed by the Walter Reed Institute here in Bethesda as a radioprotective agent. It took this product and the investigators involved in the radiotherapy program with amifostine a long journey to arrive before this committee. We believe that this SNDA shows that amifostine is safe and effective in the following radioprotective indication. To reduce the incidence of moderate to severe radiation-induced xerostomia. I thank you very much for your attention, and we are now prepared to entertain your questions.

CHAIRPERSON DUTCHER: Thank you very much. We appreciate your keeping up with the time. You did a very good job.

DR. OSTER: Thank you.

CHAIRPERSON DUTCHER: Questions for the sponsor? Yes, sir? Dr. Harwood?

DR. HARWOOD: I would like to ask the radiation oncologists who have presented why there is such a small number of definitive radiotherapy patients in this trial? As you will observe, two-thirds of the patients entered into the trial were post-operative patients. Only one-third of them were definitive radiotherapy patients. And there were only five patients with cancer of the nasopharynx -- I think 11 patients with cancer of the nasopharynx, which would be the group that I would have thought would have benefitted perhaps the most. And also, I would like to ask why it has had such relatively poor support amongst the head and neck radiation oncologists in this country?

DR. OSTER: I would like to call upon Dr. David Brizel to respond to these questions.

DR. BRIZEL: You may have to help me remember everything, but I will do my best here. I don't know if they will be in the order in which you asked, but I think that a large component of the answer to your question relates to the fact that for patients who are resected, post-operative radiotherapy is fairly standardized and it is fairly standard to give once daily radiation therapy. Now when we switch gears and go to the group of patients who are receiving definitive radiotherapy, as a head and neck radiation oncologist I am sure you are aware of the fact that there are a lot more competing entities out there. For early stage disease, once daily treatment is still fairly standard. However, when we get into the realm of more advanced stage disease, T2 disease and on up, then we start getting into competition for modalities such as hyperfractionation or accelerated fractionation or radiation with concurrent chemotherapy. And depending upon institutional policies and biases, there is direct competition for what actually compared to other cancers is a relatively small subset of patients. I mean, there are

only 40,000 head and neck cases in the United States each year, of whom roughly only 10 to 15 percent actually are enrolled on clinical trials. So I think that is why we see more post-op patients than we see definitive curative-intent patients.

Now the other question I believe that you asked related to nasopharynx cancer. And certainly with those big large fields that we put on for those patients with the whole parotid gland, including the patients in the next room sometimes, yes, that is a real problem. However, again, we are into the issue of competing protocols. And again, this trial ran from 1995 through 1997, at which time people were becoming aware, especially through the intergroup randomized trial, that there was a very definite clinical benefit associated with the use of radiation and concurrent chemotherapy, and that would have immediately and automatically excluded patients from participation in this trial.

As far as the last question that you asked me, I believe that was why was this trial so poorly supported. I don't know. Maybe I have my blinders

on. We were very enthusiastic supporters of it, and I can only speak to my institution.

DR. HARWOOD: Yes, there is a lot of -- there is a lot of patients been entered from overseas. Let me just follow-up with -- it seems to me that the main worry about this drug is this issue of tumor protection. And I am not sure that you are going to be able to determine the issue of tumor protection on patients that have had the gross bulk of their disease resected and are being given postoperative radiation. So it seems to me that that issue has to be addressed in the patients who have not had their disease resected who are being treated with definitive radiotherapy, who would have a cure range in the range of 50 percent. And it would be that group that I would be worried about the potential for tumor protection. And I would also be worried about that spreading of the survival curves occurring after 18 months. You know, we only have 18-month follow-up here. I know in post-operative patients, that may be adequate. But would you be comfortable taking a T2N0 tonsil cancer with a 50 to 75 percent chance of cure

whose salivary glands are going to be radiated and give them amifostine?

DR. OSTER: Dr. Curran?

DR. CURRAN: Yes, Dr. Harwood, with respect to the first concern you had that there was a large proportion of patients who were pot-resection receiving radiotherapy. The RTOG data base of over 500 patients that I showed on those two slides matched patients according to whether they were definitively irradiated or received post-op treatment. The two RTOG studies that we derived that data from, one was a post-operative study looking a RT plus or minus chemo, and the other was a definitive study. So when we looked at the local/regional control rates, and I showed you the one/two-year local/control tumor rates, those were based on a match of the proportionate post-operative versus definitive cases. So that even when we match with a much larger data base of American and Canadian treated patients, the local/regional control rates are basically superimposable.

And the other issue is when you talked about the issue of the bulk of tumor being removed,

those times to local failure that I showed you in the pie graph also include a majority of patients receiving postoperative radiotherapy. So still over 90 percent of the tumor failures occur within the first two years, even when a majority of patients are receiving postoperative radiation.

Now as far as the question about a T2 tonsil patient, again there is no suggestion in WR-38 that I am aware of that would suggest that there is a compromise of that definitive radiotherapy for that patient, and there is no suggestion that there is going to be some difficulty in getting treatment delivered over the tight time frame that you are going to want to deliver the treatment for that kind of patient.

DR. HARWOOD: One last question. I noticed that Dr. Antonadou's name comes up, and I noticed that she was given reference #5, and the reference #5 is anonymous. I wonder if you could share with us where Dr. Antonadou comes from and why she isn't here and where that data has been published?

DR. OSTER: The data were actually

reported at the previous ASCO meeting and we, U.S. Bioscience, obtained data printout and we also received the protocol and we wrote up a report, and this is the basis of our presentation here. Dr. Lesley Russell has analyzed the data, and she may be able to extend on this if you have any further questions.

DR. HARWOOD: Where is Dr. Antonadou?

DR. OSTER: Dr. Antonadou was trained in Paris in Guildruff, and is currently practicing in Athens.

DR. HARWOOD: Thank you very much.

DR. LIPPMAN: I wonder if you could comment on what I think is one of the major challenges in the conduct and interpretation of studies like this that the primary endpoint is obviously a subjective one and the study is unblinded. And so -- and further, since the RTOG scale is certainly well accepted in this country, I am not sure how well accepted or what kind of experience centers, for instance, in Germany, who contribute a large number of patients, have with this. And since the data are very

compelling with the issue of folks in a Grade 2 and on the face of it, Grade 1 being mild and Grade 2 being moderate, there seems like there is a lot of potential for subjectivity and potential bias unintentional. So the question is what sort of measures were taken to make sure that there was consistency in the measuring of this primary endpoint?

DR. OSTER: David?

DR. BRIZEL: I think that is a really important question, and it was a question that I asked as well as I initially became involved in this trial. At each institution at the on-study visit, if you will, each of the investigators at that institution received instruction and review as to what the RTOG scale is so that people were familiar with it. The fact that it was unblinded is actually an interesting one. Speaking as an investigator who put patients on the trial, when patients came back for their follow-up during the treatment, just being in a busy practice I tended not to look at what they had. And when I saw the study nurse in the room, I knew they had been on the study. But if I guessed as to which treatment

patients had received or not, I was wrong as often as I was right.

The other issues, though, that I think really bear on this are the fact that all of the supportive endpoints all go and point in the same direction. We saw the slide that showed the high level of correlation between late xerostomia and saliva production, late xerostomia and patient benefit questionnaire score. All of those -- I would be a lot more concerned about this if we saw xerostomia getting better and saliva getting worse or things going off in different directions. But they were all consistent. And again, the last point that I would make is that the large number of institutions that participated is actually beneficial in this respect.

CHAIRPERSON DUTCHER: Dr. Schilsky?

DR. SCHILSKY: I just want to follow-up on Dr. Lippman's question, since I am not a radiation oncologist and I have never had the opportunity to use the RTOG scale. So maybe you could explain a little bit further. Grade 1 is -- we are talking about acute xerostomia. So Grade 1 is mild mouth dryness and

Grade 2 is moderate mouth dryness. What is the difference?

DR. BRIZEL: A great question. And it is a semi-quantitative system and it is not a perfect system. The Grade 2 is moderate to severe dryness. Mild dryness, if you look in a patient's -- first of all, the patient tells you. So there is that aspect of subjectivity to it. But with mild dryness, when you look in the mouth the saliva may be a little thick. Moderate to severe dryness -- I mean, if you can find it at all -- first of all, we saw the tongue where there is nothing and the tongue blade sticks to the tongue. But if they do have any saliva, quite often it is like a rope. It is real ropey.

DR. SCHILSKY: So is this -- is this a grading scale that is determined by the examiner, by the patient, by the interaction between the two? And can you tell us, has it ever been looked at with respect to intraobserver variability? If you sent two different people or two different examiners in to examine the same patient, how consistent would the grading scale be between the two examiners?

DR. BRIZEL: Actually, I would like you to -- that is intriguing. But could you please ask the first part of the question one more time? Is it related to the -- who determines the scoring, is that it?

DR. SCHILSKY: Yes. How is the scoring actually done?

DR. BRIZEL: It is the physician who actually assigns the score, but it is based on an interaction between the physician and the patient, consisting of both historical information from the patient and the physical findings.

DR. SCHILSKY: And with respect to whether it has ever been validated across observers?

DR. OSTER: We have some experts who really -- we have some experts here with us, Dr. Kent and Dr. Le Veque, whoever wants to comment further on this. They really have expertise in this area and I think they are able to respond to this question in further detail.

DR. SCHILSKY: I would like to hear it.

DR. OSTER: Dr. Kent?

CHAIRPERSON DUTCHER: Please for the recorder state your name and affiliation.

DR. KENT: I am Kenneth Kent. I am a maxillofacial prosthodontist heading the Maxillofacial Reconstruction Center at the University of Pennsylvania Medical Center. Xerostomia is a very subjective entity. There have been numerous attempts in the literature to more accurately define objective criteria. However, for the very dry patient, there is almost no saliva to measure. When you are looking at mild versus moderate to severe, there is relative ease in determining differences between the two. I could not testify to literature citations on this. However, having spent more than 20 years educating clinicians, both medical clinicians as well as dentists in the evaluation and management of xerostomia, it is relatively easy to train someone in a standard fashion.

DR. SCHILSKY: Could I just ask one follow-up question with respect to the criteria for late xerostomia? Because looking at the criteria again, it seems to me that the grading is sort of a

two-parter. Because the grading says -- for example, Grade 1 is slight dryness of mouth/good response on stimulation. And Grade 2 is moderate dryness of mouth/poor response on stimulation. So is that an either/or or is that an and represented by that slash?

DR. OSTER: Dr. Le Veque?

DR. LE VEQUE: Could you rephrase that or repeat that for me, please?

DR. SCHILSKY: Yes. Well, I am just reading the criteria. So for late xerostomia, Grade 1, it is slight dryness of the mouth/good response on stimulation. And Grade 2 is moderate dryness of the mouth/poor response on stimulation. So since those are two different things, the question is to be Grade 2, for example, is moderate dryness of the mouth sufficient to be Grade 2, or do you have to have moderate mouth dryness and a poor response on stimulation?

DR. LE VEQUE: Well, I think that in looking at the RTOG late criteria, the inference is that in Grade 1, patients can be stimulated via gustatory stimulation or neurotransmitted

stimulations, so on and so forth. Whereas in Grade 2, those attempts at coercing saliva would likely not work. Again, as Dr. Brizel pointed out, this is a semi-quantitative scale. And as Dr. Kent also pointed out, understanding the scale is in large part being able to make the assessment visually. Because there is no attempt to actually quantify the saliva.

CHAIRPERSON DUTCHER: Excuse me, could you just give your name and affiliation for the reporter?

DR. LE VEQUE: I am sorry. I am Francis Le Veque. I am a clinical associate professor of oncology at Carmenos Cancer Institute in Detroit.

DR. SCHILSKY: This is part of my concern with respect to these data sets. This is the primary endpoint. About two-thirds of the patients were enrolled outside the U.S. I don't know how experienced those investigators were with applying the RTOG criteria. And at the great majority of centers that did participate in the study, most of the centers enrolled fewer than 10 patients on the study. And again, of course they may treat many other patients

with head and neck cancer and use the RTOG criteria routinely. I just don't know and I am trying to get a sense of -- since this is really the primary efficacy endpoint, I am trying to get a sense as to how reproducible it is and how easy it is to apply it and how likely it is that different investigators will apply it in the same way and so on.

DR. CURRAN: One comment I could make, Dr. Schilsky, is that even though on this slide of the Atlantic we call these the RTOG late toxicity scoring scale, on the other side of the Atlantic, they are called the EORTC scoring scale. So it is not unlike the performance status criteria which are sometimes called ZERBRA and sometimes ECOG and sometimes COBG. I don't know details about the German investigators, but it is my understanding they are experienced head and neck radiation oncologists. But I think that this is admittedly a semi-quantitative/semi-qualitative assessment. But even if we all sat in the room and designed the study impeccably today, we would be left with the same semi-quantitative data. And when you have -- you know, as a non-participant, when you see

such a consistent difference in observation, that is convincing to me because I don't think interobserved variation is as substantial as we suspect. Because in reality in the clinics, patients are dry or they are not dry. And the distinction is not as hard as one may think.

CHAIRPERSON DUTCHER: Dr. Margolin?

DR MARGOLIN: Along the same lines as -- not so much the quantification, which of course I share the same concerns about and also would wonder about what happened to the stimulated saliva production if it is used as part of the definition. But more importantly, on page 19 -- I don't know the slide number, but it is on the left side -- in which the .1 gram threshold was being looked at as representing a worthwhile threshold from the dental consultant, of the approximately two-thirds of patients in each randomization arm that started in the study and got to the one-year mark by the previous slides, only two-thirds of those in the amifostine arm and less than half of those in the non-amifostine arm had measurements of the unstimulated saliva

production. And even though the numbers look nice, I would be concerned about their incompleteness. I don't know -- I assume the FDA reviewer is going to do that, but I would like to know what the sponsor has to say about that.

DR. OSTER: Yes. It is a very important question and we obviously asked ourselves this question as well, and Dr. Lesley Russell has some information in this regard.

DR. RUSSELL: We obviously looked at the reasons for data being missing at this time point. And if I could have slide 130, please. As you can see, of the patients in both groups, the reasons for missing data in both treatment groups was either that the patients had died or had had progressive disease and had gone on to receive probably some further treatment and thereby were not participating in this actual assessment. The other reasons for incompleteness of data are relatively well balanced between the two treatment groups. And as patients in any trial, some are lost to follow-up. And then were just unable to perform this test.

CHAIRPERSON DUTCHER: Dr. Santana?

DR. SANTANA: My question was already answered -- asked.

CHAIRPERSON DUTCHER: Ms. Beaman?

MS. BEAMAN: Yes. In reference to one of the trials, I noticed that over half of the patients dropped out. You noted that 1 percent suffered from nausea and vomiting and then 3 percent hypertension. But would that be enough to account for over half of the dropouts in one of the trials you noted?

DR. OSTER: I am sorry, can you repeat your question, please?

MS. BEAMAN: Yes. Page number -- I'll find it. Go right ahead. I'll come back to that.

CHAIRPERSON DUTCHER: Dr. Lippman?

DR. LIPPMAN: I would just like to come back for a moment to the primary endpoint and the scale. Although Dr. Curran has sort of put this into a two point system, you are either dry or you are not dry, the system clearly states Grade 1 as being mild and Grade 2 as being moderate. The question I guess I have is was there an investigator meeting or with

the research nurses or the physicians from the different countries just with pictures like Dr. Brizel showed or some sort of sense of we know it is qualitative and we know it is not perfect, but these are -- just so we are all on the same page, these are the criteria we are using?

DR. OSTER: A very good question. Dr. Russell?

DR. RUSSELL: There was one -- there was certainly one major investigator meeting held before the start of the study for the European sites. And then prior to trial initiation at each of the sites, an initiation visit was done where the scale was gone over in great detail with both the investigators and the study coordinators.

DR. LIPPMAN: And then just a follow-up, I guess the reason why that is so important is I gather the analysis -- the FDA analysis looking at all xerostomia versus none was not as convincing as the Grade 2 or greater. So that distinction between mild and moderate becomes very important. So it is nice to know that you had that meeting.

CHAIRPERSON DUTCHER: Dr. Ozols?

DR. OZOLS: Dr. Curran, you said that the RTOG is doing some more studies with WR amifostine looking at other protective effects. If we approve it today, what will that do to those studies? Do you think they will decrease and could you tell us what those trials are?

DR. CURRAN: Yes, Dr. Ozols. The current Phase III trial within RTOG is looking at amifostine as a potential protectant against combined modality esophagitis and pneumonitis in patients with locally advanced non-small cell lung cancer. Since there is no data -- and it is a randomized study with or without amifostine. Since there is no data in this study related to that question, it is my hope that action today would actually help that study rather than hinder it. We have pilot data from two institutions on it, but no convincing data that makes me think we don't need a randomized study in that area. The other area which is under investigation is to reduce radiation related proctitis and cystitis in men receiving radiation for prostate cancer. Again,

data is only preliminary and at single institutions. So I think more work needs to be done in that area as well.

CHAIRPERSON DUTCHER: Dr. Sledge?

DR. SLEDGE: I have a number of toxicity questions. I guess the first relates to the PVQ scale. Does the PVQ scale summate benefit and toxicity or is it just benefit?

DR. OSTER: Dr. Mackowiak?

DR. MACKOWIAK: There are eight different items, as you know. Three of them were symptom questions. So to the extent they measure the patient's perception of their toxicities. And also, the other three activities of daily living and the need to use external fluid are probably the best measure of benefit. The two happen, of course, as we assessed in factor analysis, in a highly correlated way.

DR. SLEDGE: I may not have made myself clear. Do they -- are all of the factors that are being measured in essence related to head and neck type symptoms or do they include things such as the

nausea and vomiting, the hypotension -- how patients experienced those?

DR. MACKOWIAK: You are right on the first part. They are all assessed -- the head and neck symptoms related to radiation toxicities.

DR. SLEDGE: Okay. I guess the second question, the nausea and vomiting that was seen in this trial, is this -- I saw 15 patients discontinued therapy due to nausea and vomiting. Is this sort of nausea and vomiting that could be controlled with standard anti-emetics? And if so, was that allowed in the course of the study? To what degree were anti-emetics allowed?

DR. RUSSELL: We actually formally recommended that anti-emetics be administered to all patients on the amifostine group. This was not done in 31 patients, and we know that 31 patients actually received no anti-emetics. And of those 31 patients, actually 25 reported no nausea and vomiting. If we look at patients who did receive anti-emetics -- and if I could have slide 175, please -- we can see -- and I don't know if this includes the 15 patients or not,

but probably not. We can see that for those patients who received 5HT3 antagonist, they clearly did have some amelioration of nausea and vomiting. The 61 percent of those patients who received a 5HT3 antagonist as a single agent reporting no further nausea or vomiting.

DR. SLEDGE: And a third question related to toxicity. Comparing the amifostine and the control group, there is a difference of 19 patients in terms of number of patients hospitalized. But the slide said that only 6 of these were attributable to the amifostine. If that is so, what are the other 13 attributable to?

DR. RUSSELL: The hospitalizations occurring in both treatment groups were largely related to a side effect of radiation. So if they had oral mucositis or oral pharyngitis and perhaps required some feeding supplementation. And then there were other admissions for other co-morbid conditions related neither to radiation or to amifostine. And in some of these patients, they were related to their underlying tumor.

DR. SLEDGE: Help me out a little bit here. This seems to be a fairly major difference. And you are saying that amifostine markedly reduces acute oral pharyngeal toxicity. So --

DR. RUSSELL: No. Our data actually show that the incidence of both oral mucositis and pharyngitis is actually largely comparable between the two treatment groups.

DR. SLEDGE: Again, I guess I am still confused. Why the major difference then in hospitalization? I mean, it is a pretty marked difference.

DR. RUSSELL: I am at a loss to explain the difference at this point in time. What we do know is that the incidence of Grade 3 mucositis and pharyngitis was the same between the two treatment groups. And if you look at the index of weight loss that we observed, one would suggest that both nausea and vomiting and pharyngitis or mucositis wasn't having that major an impact on the patients who received amifostine, because actually they suffered less weight loss. So like you, I am a little -- I

can't explain the difference in the hospitalizations.

DR. OSTER: If I can just add to this. Obviously I think you are referring to probably the medical reviewers assessments and our assessment?

DR. SLEDGE: I am referring to your slide on page 30. It says 50 versus 31 in terms of hospitalizations.

DR. OSTER: Can we see slide 50 again? But again, I think the important point to make is obviously the amifostine-related. And we went through these case record forms page by page to verify that only six patients were hospitalized because of amifostine-related toxicities.

DR KROOK: But the issue is why -- why? Was it because of drug, nausea, vomiting?

DR. RUSSELL: Sorry, are you wanting to know the reason for the six amifostine-related --

DR KROOK: I guess I would like to know to follow-up on Dr. Sledge, why six in that arm?

DR. RUSSELL: Who were amifostine-related?

DR KROOK: Right. But was it nausea/vomiting? Was it leukopenia?

DR. RUSSELL: No, I can tell you those.

DR KROOK: Okay.

DR. RUSSELL: Two of those six patients, actually right at the end of their treatment or two or three days following their treatment, so they were still considered related toxicities, experienced a skin type reaction. One of the patients had an erythema multiform, but was also receiving carbomazepine. So the investigator was actually unable to say which definitely caused this reaction. The other four admissions, two were related to some nausea and vomiting and other general conditions. It wasn't necessarily an amifostine-related nausea. And the other two were an irradiation type reaction. Whereas you know, when you assess these patients you have a possibly/probably/definitely related problem to take, and they took the possibly related problem because they were unable to rule out the contributory effect.

DR. SLEDGE: Is this a statistically significant difference in hospitalization rates?

DR. RUSSELL: I am unable to say. I

didn't run a P value beside this.

CHAIRPERSON DUTCHER: Dr. Nerenstone?

DR. NERENSTONE: Continuing on about the safety questions, hypotension was noted to occur in 15 percent of patients despite being prehydrated. Can you tell us what hydration they received, how long, and what the overall additional time patients needed to be in the clinic to get their hydration, their anti-nausea medicine and the amifostine before their radiation therapy?

DR. BRIZEL: Typically the patients received 500 cc of hydration either orally or intravenously -- orally water and intravenously saline. And this was given rapidly. As the slides in my presentation showed, 15 to 30 minutes prior to the administration of the drug. Since the drug was given 15 to 30 minutes prior to the administration of radiation, altogether we would be looking at a period of an extra 45 to 60 minutes.

DR. NERENSTONE: With mucositis being such a problem in a lot of these patients, were a lot of these patients able to tolerate oral hydration?

That is a lot of water.

DR. BRIZEL: That is something that can evolve over time. All of these patients did have PIC lines. I cannot comment on the actual hard numbers, but what one might expect is that towards the initiation of therapy, usually with once a day radiation it takes around two weeks or so for serious mucositis to start to become manifest. And so towards the later phases of treatment, since many of these patients had their PIC lines anyway, you would be giving it IV usually.

CHAIRPERSON DUTCHER: Dr. Krook?

DR KROOK: Actually you answered part of my question that they all had PIC lines in. But going back to Mrs. Beaman's question about withdrawal, certainly those of us who have put VADs in people, we certainly have problems with PIC lines and VADs. And certainly some of the withdrawals might have been related to just this extra hour and excuse my midwestern fussing around with the extra hour or two that it takes, having met the radiation beam myself, getting through their quick was one of the nice things

to do. And certainly if I were on a treatment arm that was an hour or two before, that certainly would add to my feelings. There must have been a few withdrawals just because of that extra inconvenience. But again, I am looking at table 18 in the book that says the withdrawals are only 27.

CHAIRPERSON DUTCHER: Mr. Gruett?

MR. GRUETT: You brought up a study of 4,000 people on vomiting. You passed that by quite rapidly. Were there other studies being done on these 4,000 people and are there other studies that you have done that have shown negative information? I have had a very difficult time following all of the charts and the information because it varies so rapidly.

DR. BRIZEL: First of all, I apologize if I did not make that as clear as I should have. It was actually what I said was that within the context of this trial, there were over 4,000 administrations, doses of the drug, that were delivered to the patient, not actually that there were over 4,000 patients on the study.

MR. GRUETT: You brought that up in the

vomiting discussion when you were given your presentation.

DR. BRIZEL: In the toxicity.

MR. GRUETT: So do you have follow-up on that 4,000 and the statistical data mentioning what the limit of vomiting was?

DR. OSTER: Limit of vomiting? Sorry.

MR. GRUETT: Yes, when you brought up the 4,000 patients, you were bringing up also -- you were discussing the vomiting in this discussion. And I was wondering if you have more information or statistics confirming this.

DR. OSTER: I think Dr. Brizel only wanted to say that whereas when you look at the percentage by patient, it may occur -- you know, the percentage as it is given to you, which I think is in the range of what was it again? Nausea and vomiting?

DR. RUSSELL: It was 8 percent.

DR. OSTER: 8 percent Grade 3. But when you then look at the incidence by infusion, you find that it is less than 1 percent. In other words, those patients who have experienced nausea and vomiting in

one course, may have taken the next time a 5HT3 receptor antagonist and were then doing well and didn't have a reoccurrence of their nausea and vomiting. That is also an experience which we have from our post-marketing experience and from other trials in chemotherapy.

MR. GRUETT: Another thing that I was trying to find out in your statistics is actually what is the percentage of people that are affected by your drug. I interpolated it to be 23 percent, but I could be wrong on my understanding there.

DR. OSTER: I think the Grade 3 nausea and vomiting, as we presented, was 8 percent, right? 8 percent.

MR. GRUETT: No, in the curing of the drug -- the stopping of the dryness. Those who were successfully affected by using the drug.

DR. OSTER: Okay. Roughly you can say you have 30 percent of patients less with acute, severe xerostomia and 30 percent less with late xerostomia. So the treatment affect is roughly in this range. 30 percent less of patients have this toxicity.

CHAIRPERSON DUTCHER: Dr. Schilsky?

DR. SCHILSKY: Just a question to clarify a little bit about the PVQ. And again, I think part of what I am grappling with and I suspect others are is the fact that this was an unblinded study and we have very subjective endpoints. Therefore, it is really difficult to determine what is an actual effect of treatment and so on. So the PVQ, as best as I can see from the briefing document, is a 10-item scale. And a high number is good. That means if you are a 10, you are not having any problems. And so as I look at some of the charts that were shown, first of all the mean values rarely go below 5. And in most cases, they don't go below 6. So that would tell me that the average patient is actually not reporting much in the way of discomfort, side effect and so on, because they are at the high end of the scale.

Second of all, the difference between the curves tends to be about one unit. So I wonder if there is a way you could help us interpret the clinical significance of saying that you are having a six-day versus a seven-day.

DR. OSTER: Before I ask Dr. Mackowiak to respond to this, let me remind you, if I may, that we obviously were aware of the fact that some of the assessments may be criticized to be subjective, and that is one reason why we have in this study not only assessed xerostomia with various modalities, three, but we also assessed basically two different events, acute xerostomia and late xerostomia, at a number of defined endpoints. And I think we can say here, based on the consistency which we have through our time points and on the consistency which we have throughout measurements, and based on the correlation which we have between the measurements, that this is a set of robust data.

DR. SCHILSKY: Although it could just be a systematic bias that all goes in the same direction.

DR. OSTER: In terms of when you have a triple or when you have a P value on acute xerostomia which is below 0001 and a P value for late which is below 001, I think it is probably unlikely that you find this just by bias. But I would like to ask really John to comment on the question which you

posed.

DR. MACKOWIAK: On the bias question, in a short-term pain study if there was a one-day duration maybe I could see that happening very quickly. But you have to remember these patients received amifostine over a six-week period of time and then we are asking them 12 months or 11 months after treatment ended are you different. And as you can imagine, I don't think that a placebo effect can act that long. As far as -- are there others who -- because I will put the chart up if you would like.

DR. SCHILSKY: I think that is a valuable point you made. My other question is what -- just in general, could you give us a way of interpreting the clinical significance of a one point change in the scale? Because whatever scale you used, there rarely seems to be more than about a one point difference. So what is the clinical significance of a 7 versus a 6 or an 8 versus a 7?

DR. MACKOWIAK: I am going to put up slide 45 to show the -- the slide one before this one first, and then we can go back to that. What we saw first

was patients started -- the range was 10 to zero. I didn't show the bottom end of this scale. But patients even prior to radiation reported on average 9. Again, these are averages. Right as you would expect, we see a 3-point drop in the scores for the amifostine treated group and a 4 point drop for the control group. The clinical significance of that after radiation, that one point difference that we showed is roughly one grade level. So that change from mild to moderate. When I looked at the scores, the actual patient scores, the standard deviations around these are quite wide. And that is part of the reason why the statistical significance is affected. Because patients are affected very, very differently from patient to patient. Is there anything else I can add to that?

DR. SCHILSKY: Go to your next slide.

DR. MACKOWIAK: Oh the next slide, exactly. Thank you. Thank you for reminding me. This is the RTOG xerostomia toxicity grades. Again, we look for consecutive visits where a clinician observed a change. And I don't know which one

validates which, but my version of it, since I usually live in a PVQ world, I think they both validate each other. You can't have both of them accidentally coming to the same conclusion all the time. We saw one grade level worse or a one point decrease. One grade level improvement or one point increase.

DR. SCHILSKY: I think all you can really say from that is that they seemed to go in the same direction. I wouldn't stress the one to one because you are dealing with two very different scales.

DR. MACKOWIAK: They are very different scales.

DR. SCHILSKY: It is basically a 3-point scale versus a 10-point scale. So I might say that if somebody had a one grade level worse on the RTOG scale, I might have expected that they might have had much greater than just a .96 fall in the PVQ scale. That doesn't strike me as being much of a change at all. Whereas going from a grade 1 to a grade 2 on the RTOG scale would presumably be a much bigger change.

CHAIRPERSON DUTCHER: Dr. Simon?

DR. SIMON: I had a couple of questions.

One, what did the protocol say in terms of when the final analysis would take place in terms of the anti-tumor effect? Did it name a number of events, as protocols frequently do? Number of recurrences? What did it specify?

DR. RUSSELL: The primary protocol-defined endpoint was the incidence of local/regional control at one year. There was no predefined number of events.

DR. SIMON: And when would that final analysis take place? Did the protocol specify when the final analysis would take place? Would it take place when all patients had been followed for one year or two years?

DR. RUSSELL: When all patients had been followed for two years.

DR. SIMON: But that has not happened yet, is that right?

DR. OSTER: Well, what we have done is obviously we have analyzed the information for one year and for 18 months, and to corroborate the meaningfulness of this observation, we have looked

into the RTOG data base, which we have shown to you, which basically says two observations. Number one, 75 or 78 percent of all of the events in this patient group have been in the first year. So there is not much action in the curve after one year. Less than that happened after 18 months. And when you then compare the curve which Dr. Curran showed matched with a matching population from the RTOG experience to our patient population, it is superimposable.

DR. SIMON: But this is before the time of the protocol-specified time of final analysis. My second question --

DR. OSTER: If I can just add one thing. The primary endpoint for preservation -- for assessing preservation of anti-tumor activity was local/regional control as assessed at 12 months.

DR. SIMON: But the time of the final analysis was to be when all patients have been -- two years after the last patient entered the study, is that right?

DR. OSTER: It wasn't specified as the final analysis. We had further follow-up because --

DR. SIMON: The definitive analysis. Usually a protocol -- I would say almost every NCI-sponsored protocol would have a time of definitive analysis, either expressed in time after the last patient is randomized or expressed when a specified number of events had occurred. Did this -- okay. My second question is in your analysis of local/regional control, how are withdrawals handled? Patients who left the study for whatever reason or patients who had a distant recurrence, or patients who went off study because of toxicity or whatever. How were they handled in that analysis?

DR. RUSSELL: This analysis was taken purely on an intent to treat. So that for all patients who actually discontinued study for whatever reason, we specifically asked the investigators to continue to follow-up these patients for local control.

DR. SIMON: But I notice in your curves you show a whole bunch of censored points in the drug-treated group, very early like about two months. There is a cluster of about 6 or 7 censored

observations in the drug-treated arm that are censored at about two months. I was wondering why that is.

DR. OSTER: I am sorry, which curve is this in? Is it in the local/regional control curve?

DR. SIMON: Yes. Okay, so it looks to me like you have censored patients. It looks like you have censored some patients when they went off study perhaps for toxicity or whatever.

DR. OSTER: No, not because of toxicity.

DR. RUSSELL: No. I mean, some of these patients are lost to follow-up and we realize that. The loss to follow-up obviously does occur in this patient population. And if a patient chooses not to come back to the clinic, we have to then censor those persons.

DR. SIMON: But I don't understand why that cluster would be in the drug-treated arm and not equally distributed.

DR. OSTER: Dr. Koch?

DR. KOCH: I am not sure whether this will address your question, but you might put up slide 216.

DR. SIMON: I was looking at figure 7 in

the executive summary from the sponsor.

DR. KOCH: Yes, I know that that is the figure you are looking at. This is an attempt to provide disposition of the people that were lost. So basically in the two groups in terms of through 18 months, these are the reasons why various individuals may have been censored. Now as to which ones were censored at which time, I can't tell you that. But this is the disposition and set of reasons for the number of people who were censored prior to 18 months in the two groups.

DR. SIMON: Okay. Well, the censoring I think is of concern, particularly in a therapeutic equivalence question. My final question was for the postoperative -- the patients who received radiotherapy in the postoperative setting, what sort of increment in local control at 12 months would we believe that radiotherapy gave them? What would have been the local control rate at 12 months had they gotten surgery alone?

DR. BRIZEL: Well, typically for -- I think it is worth at this point defining the post-op

population a little bit better, because that puts the answer in its proper perspective. The post-op patients were defined as either high risk or low risk for their recurrence actually based on criteria from the randomized trial that was conducted at the M.D. Anderson Hospital and that was published about five years ago. The low-risk patients, which constituted only a small portion of the post-op patients, were those who had negative resection margins at the primary site and/or no evidence of extracapsular nodal spread in the neck dissection specimen if they had that. The high-risk patients had positive margins and/or extracapsular spread in the neck dissection. But they constitute the majority of the post-op patients and really drive the issue. I will confine my answer to that group of patients.

Patients with high-risk -- patients who were resected and have high risk pathologic characteristics, we would expect a recurrence rate of approximately 70 percent, either at the primary site or in the neck. So a local/regional control, if you will, of only around 30 percent. If we give those

patients postoperative radiotherapy, we see local/regional control rates in the realm of what was observed for the overall population in this trial.

CHAIRPERSON DUTCHER: Last question, Dr. Lippman.

DR. LIPPMAN: The proposed indication, of course, is based primarily if not entirely on one large, well-done randomized trial, and I think that is reasonable given the issues with the endpoint. I guess the question is that the proposed indication just refers to its use in radiation-induced xerostomia. And as you know, with increasing use of concomitant chemoradiotherapy, this doesn't exclude that population and there are some reasons to think that this may or may not work or it may require a higher dose and other things. It really seems like a separate question and whether the proposed indication would need to be more specific tailored to the evidence from the trial.

DR. BRIZEL: I think the trial can -- the trial addresses the patient population that was studied within the trial. And within -- as we

discussed at the beginning of the Q&A session, there are many different competing approaches for the treatment of patients with more advanced disease, as you indicate, including concurrent chemotherapy and radiotherapy. We don't know the answer to that.

DR. LIPPMAN: I guess my question is would you want to limit the proposed indication to the population you studied? Because the way it is written now, it is not clear if it includes radiation alone or radiation with chemotherapy.

CHAIRPERSON DUTCHER: We will have to talk about that when we talk with the FDA.

DR. OSTER: Okay.

CHAIRPERSON DUTCHER: We are going to take a 15-minute break and we are going to be back at 3:05.

(Whereupon, at 2:47 p.m. off the record until 3:03 p.m.)

CHAIRPERSON DUTCHER: All right, we are going to get started. Dr. Chico is going to present the FDA review.

DR. CHICO: Good afternoon, members of the advisory committee and ladies and gentlemen. I will

be presenting the FDA review for the supplemental NDA for ethyol for radiation-induced xerostomia. Our presentation is divided into two parts. We recognize that the quality of life endpoints are important, so Dr. Clara Chu, our statistician, will be presenting our methods of review and results on that.

These are the members of the FDA review team. I would like to mention special acknowledgements to the Division of Scientific Investigations and the Division of Dental and Dermatologic Drug Products reviewers for their help in our review.

The application seeks approval for ethyol for the reduction of moderate to severe radiation-induced xerostomia with primary data on a single prospective multi-center randomized Phase III trial comparing standard fractionated radiotherapy with or without ethyol in patients with head and neck cancer.

As mentioned earlier by the sponsor, the patients were randomized evenly between treatment arms. Stratification factors identified were sites of disease, clinical stage, nodal status, volume of

parotid glands for radiotherapy and type of radiation. The distribution of patients enrolled according to these stratification factors were balanced.

Type of radiation was one stratification factor which grouped patients according to the amenability of their tumors to surgery, status of tumor margins after surgery, node positivity and extension of the tumors to the neck. These factors determined the intended radiation dose. Inoperable patients received -- were prescribed 66 to 70 Gray. Postoperative high-risk patients received 60 to 66 Gray. And postoperative low-risk patients were prescribed 50 to 60.

This grouping allowed enrollment of a wide range of therapeutic and prognostic groups of patients with head and neck cancer. Administration of a wide range of doses of radiotherapy was present in the study, and we have seen this as an important prognostic factor for head and neck cancer as well as an important factor in determining the risk for late xerostomia. The FDA reviewer grouped patients according to the likelihood of experiencing late

xerostomia as in patients who received more than 45 Gray, and the likelihood of experiencing severe late xerostomia as in patients who received more than 65 Gray. This analysis was done in full recognition that factors other than total doses of radiation affect the incidence and severity of late xerostomia.

As is shown here, there is significantly more patients who received more than 65 Gray in the RT-alone arm. The difference between those receiving 45 to 65 Gray was not significant, and the overall difference in the distribution of patients was only marginally significant with a P value of .056. The effect of this distribution on the primary endpoint of late xerostomia will be shown in a later slide.

The primary endpoints related to radiation effect include acute xerostomia, acute mucositis and late xerostomia. Acute events were defined according to RTOG criteria. For acute xerostomia, those were Grade 2 or greater within 90 days of start of radiation, as well as for acute mucositis, except that those are more than Grade 3 events. For late xerostomia, this was also defined using the RTOG

criteria except that the time window use is between 9 to 12 months following radiation. This was the protocol-defined time window.

These analyses were all prospective and they were all performed in the intend to treat group, and the statistical comparison used was Fisher's exact test.

The incidence of Grade 2 acute xerostomia was significantly reduced in the patients receiving ethyol. The FDA reviewer agrees with this assessment. However, there was no difference between treatment arms in the overall incidence of acute xerostomia, which include Grade 1 and Grade 2 patients. Therefore, it appears that ethyol prevents moderate or Grade 2 acute xerostomia, but does not prevent its overall incidence.

The applicant's analysis of late xerostomia was based on a retrospective definition of events at one year or 365 plus or minus 31 days from the start of treatment. This analysis, however, showed a significant difference between treatments arms, with an advantage on patients in the ethyol

group. Although this is a close approximation of the protocol definition of 9 to 12 months following treatment, it became apparent during later discussions with the applicant that additional data would have to be submitted to the agency to determine the incidence of late xerostomia that would reflect more accurately the protocol definition. The results of the updated analysis is presented in the following slide.

The revised analysis by the FDA as shown in yellow font here were determined by determining Grade 2 or greater late xerostomia documented on visits labeled month 9 or month 11 after treatment. Our analysis was carried out on all patients treated in the study, showing findings that were similar to that of the applicants. 40 percent of the patients in the RT arm versus 24 percent of the patients in the ethyol arm reported Grade 2 or greater late xerostomia, showing a statistically significant difference in favor of ethyol, with a P value of .0015.

Please note, however, that the reminder of the applicant result is a misprint. These two columns

should be interchanged such that for patients in the RT arm, there was 34 percent greater late xerostomia and 57 percent for the amifostine arm. I am sorry, 34 percent --

DR. WILLIAMS: You have changed it on your slide.

DR. CHICO: Oh, I am sorry. This was a last minute correction, I am sorry. Although patients on the ethyol arm received a higher median dose of radiation compared to patients on the radiation arm, an advantage for ethyol was seen in each group, especially in patients who received between 45 to 65 Grays of radiation. This difference is also present in the patients who received more than 65 Gray, but it was less impressive.

Another primary endpoint of the study was acute mucositis, and the analysis of the applicant showed that there was no difference in the incidence of Grade 3 or greater acute mucositis between the treatment arms. Analysis of events between Grade 1 to Grade 4 also showed no difference.

In summary, the results of the analysis of

the primary efficacy endpoints related to radiation effects by the applicant and the FDA were generally in agreement. The robustness of a finding of a significant advantage in favor of ethyol regarding the incidence of moderate to severe acute and late xerostomia were clearly demonstrated. There was a significantly lower incidence of moderate acute xerostomia, a significantly lower incidence of moderate to severe late xerostomia, but no difference in the incidence of acute mucositis. Since the overall incidence of acute and late xerostomia were similar between treatment arms, the applicant changed the proposed indication to prevention of moderate to severe radiation-induced xerostomia in patients with head and neck cancer.

The primary efficacy endpoints related to radiotherapy involve assessments by investigators of the degree of xerostomia. On the other hand, important secondary endpoints also related to the efficacy of ethyol are measurements of saliva production and the patient benefit questionnaire. These are endpoints which provide objective

measurements of xerostomia and should reflect patients' own feedback regarding the effect of xerostomia on their quality of life respectively.

Evidence of tumor protection is provided by assessment of local/regional tumor control at one year, which was a primary endpoint, and secondary endpoints of disease-free survival and overall survival. The safety endpoint was described also.

The applicant's analysis of saliva measurements showed that there was a significant difference favoring ethyol in the unstimulated saliva production at one year using .1 gram saliva production as the cut-off to establish clinically significant saliva production. A similar analysis of stimulated saliva production at one year showed no difference between the treatment arms. A longitudinal analysis by the FDA reviewer of unstimulated saliva production done in order to establish trends over time did not support the positive findings by the applicant.

Instead of four post-radiation follow-up time points prospectively defined in the protocol, the applicant retrospectively compared measurements of

saliva only at three time points -- at baseline, at first follow-up visit, and at one year following radiation. Wilcoxon rank sum analyses of saliva collection were planned. However, details of the analysis were not specifically written in the protocol. The applicant consulted with the experts and the decision to use categorical assessments of clinically relevant levels of saliva production in the applicants analysis were decided after all the samples have been collected.

An exploratory analysis using comparison to baseline measurements was undertaken based on uncertainty from the literature in establishing a normal and abnormal range of saliva production. Tracking the change in salivary flow over time has been suggested as one method of monitoring the degree of pathology from certain disease states or the degree of damage from certain therapeutic interventions such as radiation. This analysis shows a trend of less change from baseline stimulated production for patients in the ethyol arm. For the unstimulated saliva production, there doesn't seem to be any trend

of a difference between the treatment arms. Note, however, that where there was the greatest difference between treatment arms in the stimulated production at month 11, there were only 40 percent of patients and 45 percent of patients in the amifostine arm who submitted samples for analyses.

The analysis methods and subsequent results by the applicant and the FDA were different. Analysis of saliva measurements by the applicant showed a significant difference in unstimulated saliva collection at one year, but was not supported by the FDA's longitudinal analysis of unstimulated saliva collections. Nor their analysis of the stimulated saliva collections at one year. A comparison of baseline by the FDA showed a trend toward less change for patients in the ethyol arm in the stimulated saliva collections.

To address the issue of quality of life, also different methods were once again employed by the FDA, and the difference in the methodology and results of the FDA analysis will be presented by our statistician, Dr. Clara Chu.

DR. CHU: Okay. I am Clara Chu, and I will discuss the analysis of the patient benefit questionnaire data. I will first describe the reasons why the sponsor and the FDA obtained different results after analyzing these data and then briefly discuss how the FDA analyzed these data, show some results from the analysis, and then summarize these results.

The first difference is that the sponsor and the FDA used different measures of clinical benefit. The sponsor used a mean score that was calculated from 8 questions from the questionnaire. The FDA defined three individual subscales using specific questions from the questionnaire.

The first subscale, functional well-being, was defined using two questions. One pertaining to one's ability to speak due to dryness and the other pertaining to one's ability to eat due to dryness.

The second subscale, general condition, was defined using the question corresponding to an overall feeling of dryness at rest, that is, while not eating or chewing.

Use of external aids was defined using two

questions, one corresponding to the frequency of fluid intake for eating, and the other corresponding to the frequency of fluid intake for comfort not associated with eating.

The second difference occurred in the number of data points used in the analysis. The sponsor excluded all the data points beyond the one-year follow-up visit, while the FDA included all these data points.

In the FDA analysis of the patient benefit questionnaire data, the FDA employed a pattern mixture model to investigate the impact of missing data in this analysis, that is, to determine whether or not the missing data could be considered ignorable or not treatment-related. This approach consisted of dividing patients into two groups, dropouts and completers, where dropouts were defined as those patients who dropped out before the beginning of the second year of follow-up, and completers were defined as those who remained on study through the second year of follow-up.

The table on this slide shows the number

of patients in each treatment group split by their dropout and completers status. The method used to analyze the data was a longitudinal analysis with GEE quadratic models for each of the PVQ subscales.

To give an idea of what results were obtained in this analysis, the mean scores at two-month time intervals were displayed for each treatment group and dropout and completer status group for each of the subscales. This plot displays the predicted means for each treatment group split by dropout and completer status, and it is clear that within dropouts and completers, the trends are the same between the treatment arms for functional well-being. And the same holds true for general condition and also for use of external aids.

Because the trends are about the same for treatment arms within both dropout and completer groups, the individuals were then combined and the longitudinal analysis performed. This plot shows the predicted mean scores and the observed mean scores for each two-month interval -- oh, I am sorry. This plot shows the predicted mean scores and the observed mean

scores for each two-month interval for the subscale functional well-being. The predicted means -- actually, the predicted and the observed means for ethyol are shown in red, and the radiation observed and predicted means are shown in yellow. From this plot, it is not clear whether there is a trend in favor of the ethyol arm for this particular subscale.

However, for general condition, both predicted means and the observed means do follow a trend in favor of the ethyol arm, and the same can be said of use of external aids.

To summarize the results of the PVQ analysis, first it needs to be understood that the results of this analysis should be considered descriptive and exploratory because of the difficulty in interpreting the results due to the subjective nature of the questionnaire. And also in addition, the open label trial design can result in bias of different types. Also, there is the need for adjustment of multiple comparisons.

Second, trends are in favor of the ethyol arm for the subscales general condition and use of

external aids. And lastly, it is not clear whether or not there is a trend in favor of the ethyol arm for functional well-being. I'll turn this presentation back to Dr. Chico.

DR. CHICO: Thank you, Dr. Chu. The next endpoint would be tumor control. And according to the sponsor's analysis, there is no difference between the treatment arms with respect to local/regional tumor control at one year, which was again the primary endpoint. There was also no difference between the treatment arms in the disease-free survival rate and overall survival rate. Supportive evidence from another randomized study in patients with rectal cancer reported no difference between treatment arms in overall survival rates after radiation.

At planning meetings with the applicant, the agency recommended that at least 195 local/regional failure events are needed to yield 80 percent power to exclude a hazard ratio of .7. For the analysis of local/regional failure at one year, the number of events was approximately half of that recommended. Aside from the high censoring rate,

there was also selection of a liberal lower confidence limit of 70 percent in a non-inferiority test.

Safety findings were described. Despite lower daily doses of ethyol, a significantly greater frequency of expected severe adverse events such as nausea, vomiting, fever, allergic reactions, hypotension and dizziness were observed. There was a 19 percent dropout rate in the ethyol arm. Such dropout rates raises concerns regarding the effect on efficacy results. There were more missed radiotherapy doses in the ethyol arm compared to the radiotherapy-alone arm. There were twice the number of radiation doses missed due to toxicity in the ethyol arm. However, the differences were not statistically significant. Adverse events from treatment resulted in 101 hospitalizations in the amifostine arm and 63 in the radiation-alone arm.

Regulations require that for a drug to be approved, substantial evidence of effectiveness be demonstrated through adequate and well-controlled trials. There should be substantial evidence of important clinical benefit and a tolerable toxicity

profile. There were significantly more adverse but expected events in patients treated with ethyol. There were significant numbers of dropouts, hospitalizations and missed doses. Although these were expected events, these raised concerns regarding one's ability to deliver ethyol in the treatment itself. On the other hand, one should take caution not to overinterpret toxicities in a no-treatment controlled trial. Judgment of a tolerable toxicity profile should be taken into consideration with other benefits that may be present with the treatment.

The analyses results suggesting that ethyol decreases Grade 2 to 4 late xerostomia are robust and are clearly statistically significant. Results of the analysis of saliva production in patient benefit questionnaire data may be supportive but were less impressive.

Finally, evidence to support the proposed indication comes from a single, large, adequately controlled randomized Phase III trial. In most cases, the FDA has required more than a single trial. However, in other cases, the FDA relied only on a

single adequate and well-controlled efficacy study to support approval. Generally only in cases in which a single multi-center study of excellent design providing highly reliable and statistically strong evidence of an important clinical benefit such as an effect on survival, and that a confirmatory study would have been difficult to conduct on ethical grounds.

Whether this single unblinded trial with these efficacy and safety findings is adequate to support approval is not clear, and this very important question seems ideal for consideration by the advisory committee. Thank you.

CHAIRPERSON DUTCHER: Thank you very much. Okay, questions for FDA from members of the committee? Dr. Sledge?

DR. SLEDGE: There are a couple of questions. First, getting back to the safety one. I will ask you the same question I asked the company, which is the difference in terms of hospitalizations. Is this a statistically significant difference, and do you have any sense of why there are so many more

hospitalizations with the drug?

DR. CHICO: I think it might have been my mistake not to have labeled it appropriately. The hospitalizations were based on the number of events, not on the number of patients who were hospitalized. So maybe that may explain the discrepancy in numbers. However, we did not perform a statistical analysis on the difference, but the sponsor did and they did not show a statistically significant difference.

DR. SLEDGE: Okay. With regard to the PVQ scales, my understanding from the company presentation was that this was a fairly standard scaling system, is that correct?

DR. CHICO: The patient benefit questionnaire was internally validated. This was a subject of one of our conferences with the sponsor, and they had several experts decide which were clinically significant endpoints to use as questions in the questionnaire. But it was not validated in another study.

DR. SLEDGE: So this is the only study PVQ has ever been used in?

DR. CHICO: That is what I am aware of.

DR. SLEDGE: That wasn't the impression I had.

CHAIRPERSON DUTCHER: Is that correct?

DR. MACKOWIAK: May I address that?

CHAIRPERSON DUTCHER: Yes, please.

DR. MACKOWIAK: I'm sorry, I did not mention that the PVQ was developed by the RTOG. I did notice that an instrument that was very, very similar was used in pilocarpine studies earlier, but I don't know what other studies it had been used in. I can research that and get back to you on that.

DR. SLEDGE: I guess my question related to the PVQ then is we have two different ways of analyzing the same data. Is there any reason -- I was given no reason to choose one over the other.

DR. OSTER: If I may, can I comment on this? I think what we tried to show you in our presentation was that it probably boils down to which model you are using. And the goodness of fit issue is a very important question. When you look at our model, which was chosen by our experts, you see that

basically our model fits very well and is almost superimposable in the distribution of events and measurement points to our primary endpoint, which was defined as mean score. So I think our model which we have selected indeed shows and exhibits this goodness of fit. And we were asked by the FDA to add to our initial protocol-stipulated methodology, which was analysis of mean scores, a longitudinal analysis. We chose this one. We checked this with the FDA in our statistical analysis plan, and this is how we conducted it. We also have, if you wish, the goodness of fit presentation of the model which the FDA chose, and we would be happy to share this with you.

CHAIRPERSON DUTCHER: I think the issue is whether this scale has been used in other head and neck studies. That is the question you are asking.

DR. MACKOWIAK: It has been used in other head and neck studies. And your other item -- you haven't been given information regarding which model to select. And I think that both presentations at the close show the trend favoring the amifostine arm as relates to the issue to be discussed today. Is the

information supportive? I think that is the important part to remember. There are differences in the data that are used. There are differences in the model that are used. And I don't know if we want to spend the time to go through the detail of those finer differences, but if you want us to go to the bottom line, I think both groups agree that the trend favors the amifostine arm.

CHAIRPERSON DUTCHER: Dr. Schilsky?

DR. SCHILSKY: A quick question regarding your analysis of the saliva production and your decision to do the analysis based on change from baseline. So during the sponsors presentation, they essentially offered a rebuttal to your chose method of analysis, and I wonder, having heard their rebuttal, whether you accept it or whether you still feel that the change from baseline is the appropriate analysis.

DR. CHICO: I don't think it would be fair to say which one is right and which one is wrong. The sponsor has adequately presented that certain categorical assessment of clinically significant saliva production is accepted, and it has been

accepted by our consultants from the agency. However, this analysis which I performed which is a change from baseline was based on just literature concerns that you can never establish a normal level of saliva production, and that maybe changes from baseline would be an important endpoint to look at. I recognize though that both of these analyses only reflect quantitative measurements of saliva. It doesn't reflect qualitative or functional changes in the patient.

DR. SCHILSKY: One other question about the PVQ analysis that you performed. So you decided to do an analysis differently from the analysis that the sponsor did. You decided instead of using the overall scores to break it up into these functional categories. But you didn't really tell us why you decided to do it that way and how it is that you decided which scales to group together and why you chose to group them in the ways that you did.

DR. CHICO: Dr. Chu and I talked about this when we were doing the review, and I think the reason -- the main difference in the analysis between

the sponsors and ours was we made an attempt to group the questions according to specific symptoms, which would probably be reflected more specifically in the label if approved as a significant quality of life change with the treatment, compared to just looking at the overall means of all 8 questions. This was what we thought to be clearer and clinically significant.

MR. GRUETT: Did you do any analysis on tumors that may possibly exist with in the saliva glands?

DR. CHICO: No, we did not. No. Maybe --

DR. BRIZEL: The tumors -- first of all, patients had to have squamous cell carcinoma, which in and of itself is an exceedingly unusual histology to arise in the saliva gland. And secondly, because of tumor protection concerns, patients who presented with primary salivary gland tumors were specifically excluded from eligibility in this trial.

MR. GRUETT: I have one more question. The confusion between the analysis of the data, was this called out in the protocol?

DR. CHICO: Which analyses? Would you

please specify?

MR. GRUETT: What we have been discussing, the difference in your analysis of information and their analysis. Was there a variation in the protocol?

DR. CHICO: I think there is a table -- if you have a copy of my review -- on page 18, which summarizes the analyses that were intended and the analyses that were actually done. For different endpoints -- endpoints like late xerostomia, acute xerostomia and acute mucositis, the analyses of those were all prospectively defined in the protocol. It was just in the analyses of the secondary endpoints where there was a little bit of contention between the sponsors and the FDA.

CHAIRPERSON DUTCHER: Dr. Ozols?

DR. OZOLS: I am most concerned about your conclusion that the secondary endpoint of tumor control is premature to make a conclusion that there is no potential decrease in tumor control.

DR. CHICO: We had recommended during our meetings with the sponsor that ideally they should

probably have at least 195 events to document adequately the curve of local/regional tumor control. However, there were only about 100 events reported at one year to document local/regional tumor control. So that is the reason why we thought this was probably premature.

CHAIRPERSON DUTCHER: Dr. Margolin?

DR MARGOLIN: Just a clarification to help me understand, Dr. Chu, your analysis. In these graphs that you showed us, I assume that what you were really doing based on this predicted and observed model is trying to demonstrate that as time goes on the number of dropouts versus completers gives you much more variability and a less tight curve to compare between the groups. Is that correct?

DR. CHU: That is correct. As time goes by, there are very few events, and that is why those lines are going all over the place.

DR MARGOLIN: Thank you.

CHAIRPERSON DUTCHER: Dr. Simon?

DR. SIMON: Could you review how many patients had saliva samples available at the 12-month

analysis?

DR. CHICO: Before I answer that, I would like to just emphasize that the time windows that were used in the collection of saliva had very great variability. The time labeled as month 1 was collection between month 0 to month 3, month 5 were collections between month 3 to month 6, while month 11 were collections between month 6 to month 15, which is a 9-month time span. But to answer your question, on the events that were labeled as month 11, for the stimulated -- let me just flash the slide. If you would just refer to page 17, which is the graph of change from baseline saliva measurements. At month 11 for the stimulated production between treatment arms, in the amifostine arm, there was only 45 percent of patients who provided samples, while 40 percent of patients provided samples in their radiotherapy arm.

DR. SIMON: So it is like 45 percent?

DR. CHICO: Over here.

DR. SIMON: Could you focus that? What about earlier on? So what about at an earlier time?

DR. CHICO: All right. For month 1 for

the stimulated saliva production, there were 65 percent of the patients in the amifostine arm provided samples while there were 71 percent of the patients in the radiation arm that provided samples.

DR. SIMON: So is it your impression that there is a large difference between whether you do change from baseline or just absolute value of the --

DR. CHICO: It is my impression that between treatment arms in the stimulated collections, there seems to be a lesser degree of change from baseline. Notice that we did not perform any statistical analysis because we want this to be plainly a description and exploratory. And because of mainly the concern was the large number of dropouts in month 11, and you have to consider for multiple analysis endpoints if you want to do a -- try to confirm it by doing statistical analyses.

DR. SIMON: This was less true for the PVQ issue in terms of missing data?

DR. CHICO: There was also a large number of missing data in the patient benefit questionnaire, and I think Dr. Chu can expound on that some more.

DR. SIMON: I mean, for example, what percentage of the patients had questionnaires around one year?

DR. CHU: We don't have the slide with the actual frequency counts at one year, but I can refer you to my review. There is a frequency plot in the review.

DR. SIMON: Do you remember just approximately what it was?

CHAIRPERSON DUTCHER: Identify yourself.

DR. CHEN: This is Gang Chen, Biometrics team leader, FDA. So the frequency at about 12 months is around 80 percent for both groups.

DR. SIMON: One other question I had was there was this difference in radiation dose delivered. Was it your impression that that was attributable to missed doses because of toxicity of the drug and dropouts?

DR. CHICO: In the first place, the patients were randomized with prescribed doses. And some of these patients, especially those who were not operable, were prescribed 66 to 70 Grays of radiation

when starting treatment.

DR. SIMON: But it was balanced with regard to --

DR. CHICO: It was. That is right. Let me show --

DR. SIMON: But what they actually received was not very well balanced.

DR. CHICO: That is true. Let me just --

DR. HARWOOD: The difference in dose between the two arms I think was very small and really not expected to have a significant effect.

DR. SIMON: Well, it is 56 percent of the radiation only had over 65 Gray, and it was 43 percent in the drug plus radiation.

DR. WILLIAMS: Dr. Simon, the sponsor did a frequency distribution I think on page 13 of their slides, and I think that our analysis may be somewhat an artifact of where we just happened to pick. We picked it for I think a physiological reason. But if you look at the frequency distribution, I think we might have just picked a place where it was trending in the other direction. So I am not sure that one can

say that overall it was unbalanced. It was in that particular analysis.

DR. CHICO: And I think that was clearly stated that our overall comparisons did not show a significant difference with a P value of .056.

DR. SIMON: Well, okay. Just for my own knowledge, is this stimulated saliva production more important medically than unstimulated?

DR. CHICO: From -- maybe our experts can answer that.

CHAIRPERSON DUTCHER: Dr. Le Veque, can you identify yourself?

DR. LE VEQUE: Thank you. Let me answer your question first about the stimulated saliva, and maybe I can in one presentation also address the problem of differentiating between baseline and residual saliva. This study, unlike other studies that have addressed salivary gland dysfunction from radiation therapy, is a preventive study in the sense that -- which in toxicity I believe is a gold standard, to prevent the toxicity rather than to treat the toxicity. In this instance, as has been pointed

out I think very carefully by the sponsor, is that the end of this study, patients who were responders, patients who were in the amifostine arm, had residual saliva. They would go to bed at night and wake up with it in the morning without stimulation.

The question of stimulation is an interesting one but not really meaningful in a scientific sense because the instances where you stimulate saliva -- saliva can be stimulated many ways. In the pilocarpine studies, the whole concept of pilocarpine is one of stimulation. In the pilocarpine studies, the patients who were on pilocarpine didn't do very well with stimulated saliva because it was already -- the glands were already being stimulated. In this instance, the type of stimulation that was used in this study I believe was somewhat flawed and I think the sponsor will acknowledge this. The stimulation collection was done hard on the heels of the unstimulated collection in glands that were pretty beat-up from the standpoint of obvious production of salivary flow. So right after the unstimulated collection, the stimulated collection

occurred, and the methodology used was a parafilm chewing, which is mechanical stimulation, which is the poorest of all stimulatory mechanisms in this mode.

So I would offer the comment that the baseline issue is, from the prospective of a clinician who works with this patient population, is not meaningful. What is meaningful is we don't zero out in these patients. We have something left. And not only do we have something left, we see a slight improvement over a period of time.

DR. SIMON: Just to clarify. The FDA's analysis of the unstimulated -- change from baseline in unstimulated saliva showed no difference.

DR. CHICO: The FDA analysis of the unstimulated saliva showed that the trends between treatment arms were similar.

DR. SIMON: Right.

DR. WILLIAMS: I think a good take-home message for future trials is how important it is to be very specific in specifying your analyses. Because we can never say what is the right thing -- we can say maybe we would prefer one or the other, but we can

never be assured that something is not data driven unless we have the analysis prior to the data. So I think it is very important to specify in advance your planned analysis.

CHAIRPERSON DUTCHER: Dr. Chen?

DR. CHEN: I have a few comments regarding the statistical difference or the statistical analyses performed by the sponsor and by the FDA. The first issue I want to discuss is about Dr. Gary Koch's comments on the higher censoring rate. Actually, in our review -- in the FDA review, we did mention about the higher censoring rate. The censoring here we defined it as actually censoring patients consists of those patients who were lost to follow-up and those patients who were not events at the study cut-off date. So that is why we see the very high censoring rate. In other words, actually we are talking about the lower event rate. The second issue regarding the measure -- the health ratio we used for the time to event analyses, I think this is very important and also it is a well-known measure used for time to event analyses. And I think it is appropriate.

The third issue is regarding the different models we used for PVQ analysis. We had a meeting with the sponsor and we did agree that the longitudinal analysis was a mixed model that the sponsor proposed. The issue is not what model we should use. The question is how should we interpret the results for those model-based analyses. So actually I could let Dr. Chu present our results and the interpretation. So those are my comments.

CHAIRPERSON DUTCHER: Other questions for FDA? Any other clarifications people need before we discuss? Dr. Simon?

DR. SIMON: Just in terms of -- could you summarize your analysis of the PVQ data? The conclusions you reached from the PVQ data?

DR. CHU: Well, one, the first bullet in the summary probably being the most important is that the results that we arrived at from the analysis is that they should be looked at as being descriptive and exploratory. This would be because of like the subjective nature of the questionnaire and also due to the fact that we have an open label trial design in

the study and also the fact that we need to adjust for multiple comparisons. The other bullets are simply basically just saying that the trends are in favor of ethyol simply by eyeballing the plots.

DR. SIMON: Is that as strong a statement as you can make is that the trends are in favor of one group?

DR. CHU: Yes. That is about as strong a statement as we can make.

CHAIRPERSON DUTCHER: Other questions? Yes, sir.

DR. WASSERMAN: My name is Todd Wasserman. I was just going to ask Dr. Chico, as I understand it, the change in baseline analysis on the salivary function was just something you decided to use because you had no other defined way of analyzing the data?

DR. CHICO: No. This was something that I used based on literature which stated that that is one possible way of tracking changes in saliva production from disease or from therapy. It is not my own.

CHAIRPERSON DUTCHER: Thank you very much.