CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:NDA 20221/S012

STATISTICAL REVIEW(S)

Statistical Review and Evaluation

JUN 14 1999

NDA #:	20-221 / S-012
Sponsor:	U.S. Bioscience, Inc.
Name of Drug:	Ethyol (amifostine) for injection
Indication:	Treatment of radiation-induced xerostomia in head and neck cancer
Documents Reviewed:	Volumes 1, 10, 11, 12, 13, 14, 16, 33 dated 12/24/98
Medical Reviewer:	Isagani Chico, M.D.

This review consists of six sections. The first section provides some brief background information, and describes major statistical issues. The second section provides a description of the study. In section three the primary efficacy results are presented. In section four the secondary efficacy analyses and results are presented. In section five an overall summary is provided, and in the last section conclusions and recommendations are given. Some sponsor's Tables and Figures are attached in the Appendix.

I. Background and Major Statistical Issues

Ethyol (amisfostine) has been evaluated for its ability to protect against toxicities associated with radiation treatment. In this supplemental NDA, the sponsor seeks approval of Ethyol for pretreatment with standard fractionated therapy for the indication of xerostomia in patients with head and neck cancer. The sponsor's submission includes one pivotal randomized phase III trial (WR-0038) of Ethyol ± radiation in patients with head and neck cancer and two additional randomized studies conducted by independent investigators. Reports from two additional controlled clinical trials demonstrating Ethyol's ability to protect epithelial-like tissues other than salivary glands from radiation-induced toxicites are also included. This statistical review will only focus on study WR-0038. Several major statistical issues are identified by this reviewer. They are

1. Inappropriate Type I error adjustment for multiple endpoints and interim analyses.
2. An inappropriate selection of a liberal lower confidence limit in a non-inferiority test and high censoring rate in evaluation of local regional tumor control.
3. Inappropriate analyses for QOL data.
4. Retrospective selection of time points in comparison of the treatment arms.

II. Description of the Study

Objective

The objective of this study was to determine if pre-treatment with Ethyol decreases the incidence of oral radiation toxicities without decreasing the anti-tumor efficacy of radiation.

Design

WR-0038 is a phase III, open-label, multi-center, randomized, parallel group trial comparing two treatment arms:

1. Ethyol (200 mg/m^2 i.v. 15-30 minutes prior to radiation therapy) + radiation therapy (1.8-2.0 Gy given 5 days/week for 6-7 weeks)
2. radiation therapy alone (1.8-2.0 Gy given 5 days/week for 6-7 weeks)

for treatment of patients with head and neck cancer. Prior to randomization via a dynamic allocation scheme, patients were stratified on the following factors:

1. site of disease (oropharynx vs. nasopharynx vs. oral cavity vs. larynx)
2. nodal status (N_0 vs. N_+)
3. Karnofsky performance status (100, 90, 80 vs. 70, 60)
4. type of radiotherapy (post-operative low risk patients vs. post-operative high risk patients vs. definitive patients)
5. treatment center (40 centers).

Patients were assessed weekly. First follow-up was one month after completion of protocol therapy, and thereafter, every 2 months for an additional 10 months in the first year. In the second year, follow-up was every 6 months. Patient benefit questionnaires were filled out at baseline, weekly while on therapy and at each follow-up visit. Salivary sampling took place at the following follow-up visits: 1, 5, 11, 17, and 23 months after radiation therapy.

Patient Population (Protocol)

The protocol specified a sample size of 250 patients evaluable for oral radiation reactions and local regional tumor control at the 9-12 month time assessment with 125 randomized into each treatment arm. "Statistical power for detecting differences in the reactions and observing a lower limit of one-sided 95% confidence (LCL) no less than 0.7 for the ratio of tumor control rates between treatment groups are as follows:

Statistical Power for Detecting Differences Between Treatment Groups or Demonstrating Equivalency in Antitumor Efficacy				
Parameter	Hypotheses		Power (analyzed separately)	Power (multiple endpoints adjusted)
	H_0	H_a		
≥Grade 2 Xerostomia	80%	55%	99%	≥98%
≥Grade 3 Mucositis	50%	30%	90%	≥84%
Late Effect Xerostomia	55%	35%	88%*	not applicable
Local Regional Tumor Control	55%	55%	92%**	LCL >0.7 (interim analysis) LCL>0.71 (final analysis)

* An interim analysis with α=0.01 and final analysis with α=0.048 are planned. Statistical power is 46% in the interim analysis and 88% in the study.

** An interim analysis for LCL>0.7 and final analysis for LCL>0.71 are planned. Statistical power for observing LCL above the critical values is 79% in the interim analysis and 92% in the study.

Reviewer's Comments:

1. A protocol-specified interim analysis was conducted for late-effect xerostomia and antitumor efficacy when 160 evaluable patients had been assessed at the follow-up visit in the 9-12 month interval. No early stopping was planned for this study (stated in the clinical report not the protocol). The results of the interim analysis were not included in this submission. It is not clear to this reviewer what the purpose of the interim analysis was if no early stopping was planned.

2. The selection of the lower limit of a one-sided 95% confidence interval of 0.7 for local regional tumor control is very liberal. To demonstrate non-inferiority, the Oncological Division requires a lower limit of a 95% confidence interval to be at least 0.8.

Efficacy Endpoints

The primary efficacy endpoints include the following: incidence of grade 2 or higher acute xerostomia, grade 3 or higher acute mucositis, grade 2 or higher late-effect xerostomia, and local regional tumor control rates at 12 months. Acute oral radiation reactions were those occurring during treatment and up to 90 days from the start of therapy. The severity of these oral radiation reactions was determined by the RTOG Acute Radiation Morbidity Scoring Criteria. The late effect xerostomia was defined as occurring 9-12 months following radiation therapy. Severity of late effect xerostomia was determined by the RTOG Late Radiation Morbidity Scoring Criteria. Anti-tumor efficacy was evaluated using local regional tumor control at 12 months as the primary endpoint. Local regional control was defined as the following: "a patient will be assigned as a success under the category of tumor control if for that patient there is no evidence of local or regional recurrence."

The secondary efficacy endpoints include disease-free and overall survival assessed at the two year follow-up visit, scores calculated from a patient benefit questionnaire on radiation toxicity effects, and the measurement of whole saliva production.

Statistical Methods (Protocol)

1. Pearson chi-square tests were used to analyze rates or proportions (acute xerostomia, acute mucositis, late effect xerostomia, local regional tumor control). The method developed by Westfall and Young (1989) was used to adjust p-values for two endpoints: acute xerostomia, acute mucositis. In addition, for 12 month local regional tumor control rates, lower limits of a one-sided 95% confidence interval were used to assess statistical equivalency between the treatment groups.

2. The logrank and Kaplan Meier procedures were used to analyze time to local regional tumor failure, disease progression, and overall survival. Cox proportional hazards models were used to obtain hazard ratios.

3. The sponsor proposed a missing data imputation strategy. Then total scores and change from baseline would be analyzed for treatment differences. Also, a repeated measures analysis was planned.

Study Patients

All patients who met eligibility criteria and received at least 40 Gy of radiation therapy were evaluated for efficacy. Between 10/31/95 and 8/31/97, 315 patients with head and neck cancer

were stratified by site of disease, nodal status, Karnofsky performance status and type of radiation, and randomized into the two treatment arms. The Ethyol + radiation therapy arm consisted of 157 patients and the radiation alone therapy arm consisted of 158 patients. There were 12 patients (7 in Ethyol + radiation, 5 in radiation alone) who were never treated. Consequently, they were excluded from all safety and efficacy analyses. The intent-to-treat population consisted of 150 patients in the Ethyol + radiation arm and 153 in the radiation alone arm.

The sponsor also defined an evaluable patient subgroup consisting of 278 patients: 127 in the Ethyol + radiation arm and 151 in the radiation alone arm. In the Ethyol + radiation arm 22 patients stopped Ethyol and 1 had < 75% of each parotid gland in the treatment fields. In the radiation alone arm 1 patient had < 75% of each gland in the treatment fields.

III. Sponsor's Primary Efficacy Results and Reviewer's Comments

Reviewer's comment: There are 4 primary efficacy endpoints defined in the protocol. Two have been treated as multiple endpoints: acute xerostomia and acute mucositis, and two have been evaluated in an interim analysis: late xerostomia and local regional tumor control. For late xerostomia the Type I error is specified to be 0.01 in the interim analysis and 0.048 in the final analysis. For local regional tumor control the lower limit of a one-sided 95% confidence interval will be calculated and compared to a critical value of 0.70 in the interim analysis and 0.71 in the final analysis. The sponsor has not appropriately adjusted for Type I error inflation in this composite definition. Given the inherent multiplicity of this composite definition, the Type I error should be more stringent than that which the sponsor has specified.

The two treatment groups are well-balanced with respect to baseline demographic, tumor, and laboratory characteristics. The following are the sponsor's results for the four primary endpoints based on the final data analyses.

Acute Xerostomia

The incidence of grade 2 or higher acute xerostomia was analyzed using Fisher's exact test. The test revealed a highly statistically significant ($p<0.0001$) difference in the two treatment arms.

Acute Mucositis

The incidence of grade 3 or higher acute mucositis was analyzed using Fisher's exact test. The test revealed a nonsignificant difference ($p=0.4648$) in the two treatment arms.

Late Xerostomia

The incidence of grade 2 or higher late xerostomia was analyzed using Fisher's exact test. The test revealed a statistically significant ($p=0.0019$) difference in the two treatment arms.

Anti-tumor Efficacy

Reviewer's Comment: It is not sufficient to just compare the proportions of local regional tumor control (LRC) at 12 months between treatment arms. To put this simple rate comparison in context, a time to event analysis is essential. It is extremely important to consider the degree and pattern of censoring. The sponsor's rate comparison results are provided in Table 1:

Reviewer's Table 1. Sponsor's Local Regional Control Analysis

		Eth + Rad	Rad	p-value
12 month data	LRC	72%	71%	1.00
	LRC ratio	1.008		
	Two-sided 95% CI	(0.864,1.175)		
	Lower limit of 95% one-sided CI	0.886		
18 month data	LRC	61%	64%	0.70
	LRC ratio	0.956		
	Two-sided 95% CI	(0.792,1.155)		
	Lower limit of 95% one-sided CI	0.816		

The sponsor performed a time to local regional failure (LRF) analysis, where local regional failure was calculated from therapy start date and defined as follows: "1) documented disease progression with positive local tumor status, 2) documented disease progression with no information on tumor status in database, 3) patients who entered the study with disease and had never experienced NED (failure on day 1), and 4) patients who entered the study with disease and never experienced NED before an additional surgery to remove tumor (failure on day 1)." For patients who did not experience death with disease or LRF, time to local regional failure was censored on the latest date with data. The sponsor's results are provided in Table 2 below:

Reviewer's Table 2. Sponsor's Time to LRF Analysis

	Eth + Rad (# censored)	Rad (# censored)	Hazard Ratio (2-sided 95% CI)	Lower Limit of 95% 1-sided CI
12 month data	150 (106)	153 (106)	1.013 (0.671,1.530)	0.72
18 month data	150 (98)	153 (102)	0.946 (0.643,1.392)	0.68

The estimated hazard ratios are close to 1 based on either 12 month data or 18 month data, which were submitted by the sponsor on 4/30/99. The lower bound of the 95% one-sided confidence interval for the 12 month data is slightly higher than 0.71; however, the lower bound of the 95% one-sided confidence interval for the 18 month data is slightly lower than 0.71. This reviewer believes that the instability of these lower 95% one-sided confidence limits is due to an inadequate number of events (high censoring). This issue was addressed in the FDA-sponsor teleconference on 2/3/97. The consulting statistician at this conference stated that one would need 195 failure events in a sample size of 300 patients to achieve a lower 95% one-sided confidence limit greater than 0.71 with 80% power.

IV. Secondary Efficacy Results

Disease-free and overall survival

Disease-free survival (DFS) and overall survival (OS) were secondary endpoints for evaluating anti-tumor efficacy. Disease-free survival was calculated from the start of therapy until documentation of disease and/or death. Patients were censored if they had no documentation of disease progression. Overall survival was calculated from the start of therapy until

documentation of death. Patients who were alive were censored at the latest date. The sponsor's results are provided in Table 3 and in the sponsor's Figures 8, 9, A and B (see appendix):

Reviewer's Table 3. Sponsor's Disease-free and Overall Survival Analyses

		Eth + Rad (# censored)	Rad (# censored)	Hazard Ratio (2-sided 95% CI)
12 month data	DFS	150 (101)	153 (100)	1.035 (0.702,1.528)
12 month data	OS	150 (125)	153 (113)	1.585 (0.961, 2.613)
18 month data	DFS	150 (92)	153 (94)	0.99 (0.689, 1.423)
18 month data	OS	150 (116)	150 (108)	1.351 (0.865, 2.109)

* Hazard ratio: Rad/Eth+Rad, hazard ratio > 1 indicates in favor of Eth+Rad

The results in the above table show that based on either 12 month or 18 month data, no significant differences in DFS or OS were observed between the two treatment groups. Patients treated with Ethyol + radiation demonstrated a better trend in survival although it is not significant. The insignificant findings on survival may be due to an inadequate number of events. This is a concern as it bears on the anti-tumor efficacy question.

Patient Benefit Questionnaire

The sponsor analyzed the patient benefit questionnaire data by comparing mean scores between the two treatment groups at all time points and reporting p-values based on t-tests. The means were obtained by adding the scores from all questions answered (except question 3) and dividing by the number of questions answered. At least 6 of 8 questions had to be answered in order to calculate the mean, otherwise it was considered missing. The primary comparisons were undertaken at the end of therapy and at the one year follow-up visit. Similar analyses were performed on changes from baseline. These results are shown in the sponsor's Tables 15 and 16 in the appendix.

Reviewer's comment: P-values provided in the sponsor's Tables 15 and 16 should be interpreted with caution. An appropriate adjustment for multiple comparisons should be taken into consideration. In addition, this approach ignores the impact of missing data, which may very likely be informative. In such a case, bias will surely result.

Although not specified in the protocol, the sponsor performed a longitudinal data analysis using mixed models with spline functions. The analysis was performed on the mean score. The sponsor also investigated the missingness issue by dividing patients into non-completers and completers. The non-completers were those patients who had no QOL data beyond the month 5 follow-up visit, and completers were those patients who had at least one data point beyond the month 5 follow-up visit. There were 228 (113 - Ethyol + radiation, 115 – radiation alone) patients defined as completers and 73 (36 – Ethyol + radiation, 37 – radiation alone) patients defined as non-completers. The results of this analysis are shown in the sponsor's Figures 2 and 3 in the appendix. This analysis replaced the repeated measures analysis that was planned in the protocol.

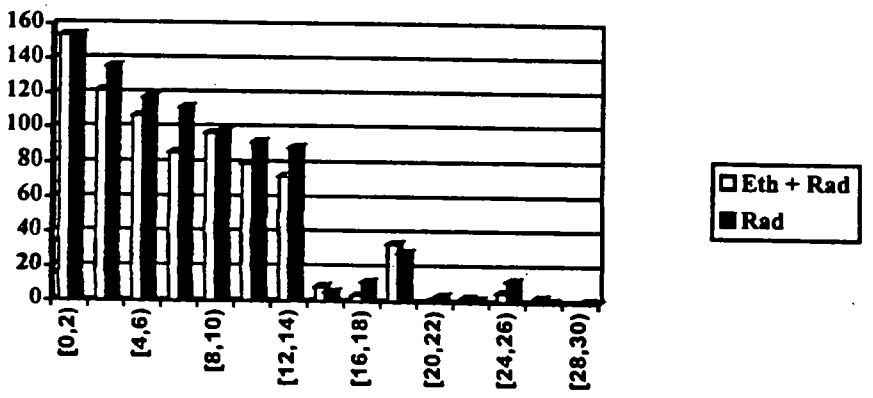
Reviewer's comment: Computing the mean scores presumes that each question has equal weight in determining clinical benefit, which may be questionable.

The medical officer believed specific questions in the questionnaire had greater clinical relevance. The 8 questions in the questionnaire were reduced by the medical officer to 3, which were considered to be the most clinically relevant. These QOL parameters were "functional well-being", "general condition", and "use of external aids". The parameter "functional well-being" was defined as the minimum score of question 4 "Speaking: Please rate the difficulty you experience in your ability to talk due to dryness" and question 5 "Eating: Rate the difficulty you experience in your ability to chew and/or swallow due to dryness". The parameter "general condition" was the same as question 1 "Dryness: Please rate the dryness of your mouth at rest (that is while not eating or chewing)". The parameter "use of external aids" was defined as the minimum score of question 6 "Rate the frequency of fluid intake to assist in eating" and question 7 "Oral comfort aids: (Saliva substitute, Water, Candy) Frequency of fluid intake required for comfort not associated with eating". For all QOL parameters, the higher the score reported, the "better" the patient felt. For example, in functional well-being a higher score implies less difficulty in speaking and eating. In general condition a higher score implies less dryness at rest, and in use of external aids a higher score implies less frequency in the use of external aids to assist in eating and for comfort not associated with eating.

If a patient visited more than once within a week for treatment, the mean, minimum, and maximum scores within the week were used in the analyses. The estimates obtained from these exploratory analyses were similar. The results are shown in reviewer's Tables A1-A3 and B1-B3 (see appendix). Because of this similarity, only the results from the analyses of the mean scores are reported.

To analyze the Reviewer's QOL data, this reviewer first examined whether the dropout rates were similar in the two arms. Reviewer's Figure 1 shows the frequencies of patients at two-month intervals starting from baseline. From Figure 1 it is evident that throughout the treatment period and during the first year of follow-up, both arms display similar rates of attrition. Attrition is especially high in the second year of follow-up for both arms.

Reviewer's Figure 1. Number of patients over time



Due to the similar dropout patterns, this reviewer further investigated the time trends of the QOL parameters for the "completers" and "dropouts" in the two treatment groups. This reviewer employed a growth curve analysis, using month as the unit of time. An estimated time trend describes the effect of treatment over the study period. This reviewer also analyzed the QOL data by using a GEE quadratic model. The generalized estimating equation (GEE) approach was developed to cope with the potential problem of informative correlation among observations per subject. An advantage of a GEE approach is that it is not necessary to specify the correct correlation structure in advance. Using the idea of M-estimation theory (Huber, 1967; White, 1982; Liang and Zeger, 1986), the solution to the (potentially mis-specified) covariance matrix is consistent. Also, M-estimation protects against under-estimation of the covariance matrix by introducing "sandwich" estimators. Therefore, we have some assurance of a variance estimate that is robust.

The models assume an "ignorable" missing mechanism and include both linear and quadratic terms. If the test statistics of the quadratic and/or linear terms were not significant, then those terms were dropped. Reviewer's Tables 4 and 5 summarize the results of the longitudinal analyses for the two treatment arms.

Reviewer's Table 4. QOL Data Analyses: Ethyol + Radiation / All Patients

Parameter	Estimate	Standard Error *	P-value
functional well-being			
intercept	5.561	0.181	< 0.0001
linear	-0.106	0.045	0.020
quadratic	0.006	0.003	0.012
general condition			
intercept	7.372	0.162	< 0.0001
linear	-0.286	0.042	< 0.0001
quadratic	0.013	0.002	< 0.0001
use of external aid			
intercept	6.538	0.185	< 0.0001
linear	-0.274	0.048	< 0.0001
quadratic	0.011	0.002	< 0.0001

* Robust standard errors are provided.

Reviewer's Table 5. QOL Data Analyses: Radiation / All Patients

Parameter	Estimate	Standard Error *	P-value
functional well-being			
intercept	5.259	0.186	< 0.0001
linear	-0.032	0.040	0.418
quadratic	0.003	0.002	0.123
general condition			
intercept	6.812	0.171	< 0.0001
linear	-0.413	0.041	< 0.0001
quadratic	0.017	0.002	< 0.0001
use of external aid			
intercept	6.132	0.177	< 0.0001
linear	-0.346	0.042	< 0.0001
quadratic	0.013	0.002	< 0.0001

* Robust standard errors are provided.

The results from Reviewer's Tables 4 and 5 show that for the Ethyol + radiation arm, there is a statistically significant quadratic time trend for all 3 QOL parameters. In the radiation alone arm, there is a statistically significant quadratic time trend for both "general condition" and "use of external aids." For "functional well-being" the time trend remains constant. Reviewer's Figure 2 displays the estimated time profiles for each arm and each QOL parameter.

Next, a pattern mixture model (Little, 1993 and 1995) was employed to assess the type of missing data mechanism. Patients are divided into two groups: those who dropped out before the beginning of the second year of follow-up and those who remained on study through the second year of follow-up. The former group is classified as "dropouts" and the latter "completers". This cutoff was selected by the medical officer because both treatment arms experienced a 50% dropout rate in the beginning of the second year of follow-up. There are 209 patients (105 – Ethyol + radiation, 104 – radiation alone) denoted as dropouts and 100 patients (49 – Ethyol + radiation, 51 – radiation alone) denoted as completers. In Reviewer's Tables 6 and 7 the results are presented for dropouts. For both arms, there are significant linear and quadratic time trends for all 3 QOL parameters.

Reviewer's Table 6. QOL Analyses for Dropouts: Ethyol + Radiation Arm

Parameter	Estimate	Standard Error *	P-value
Functional well-being			
Intercept	5.783	0.237	< 0.0001
Linear	-0.525	0.101	< 0.0001
Quadratic	0.043	0.008	< 0.0001
General condition			
Intercept	7.714	0.189	< 0.0001
Linear	-0.612	0.084	< 0.0001
Quadratic	0.043	0.007	< 0.0001
use of external aid			
intercept	7.021	0.229	< 0.0001
linear	-0.711	0.100	< 0.0001
quadratic	0.050	0.008	< 0.0001

* Robust standard errors are provided.

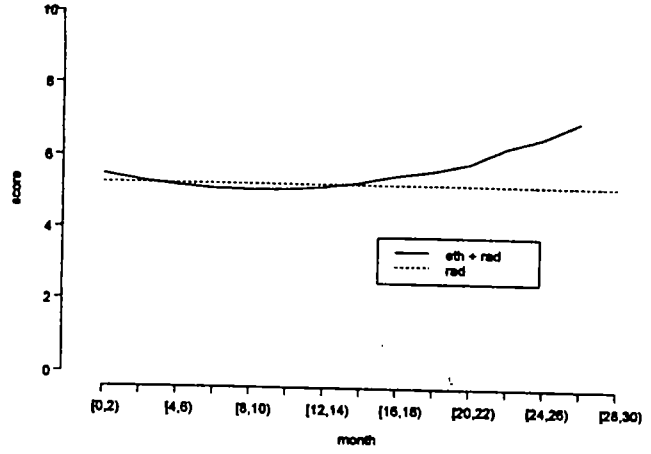
Reviewer's Table 7. QOL Analyses for Dropouts: Radiation Arm

Parameter	Estimate	Standard Error *	P-value
functional well-being			
intercept	5.374	0.257	< 0.0001
linear	-0.267	0.097	0.006
quadratic	0.026	0.007	0.0003
general condition			
intercept	7.295	0.197	< 0.0001
linear	-0.930	0.085	< 0.0001
quadratic	0.066	0.007	< 0.0001
use of external aid			
intercept	6.508	0.228	< 0.0001
linear	-0.759	0.085	< 0.0001
quadratic	0.055	0.007	< 0.0001

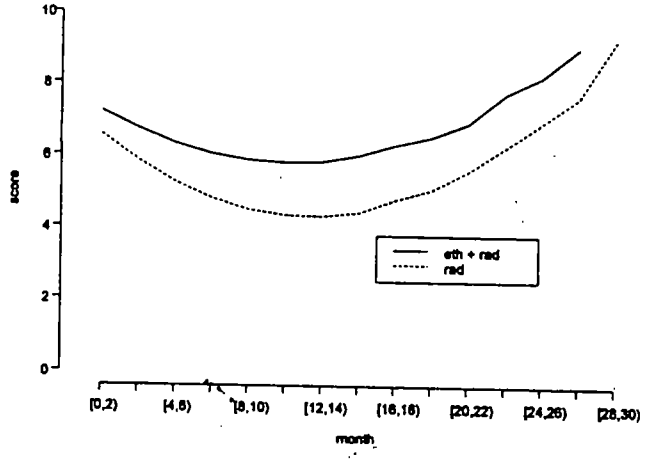
* Robust standard errors are provided.

Reviewer's Figure 2. Time trends for QOL parameters / All Patients

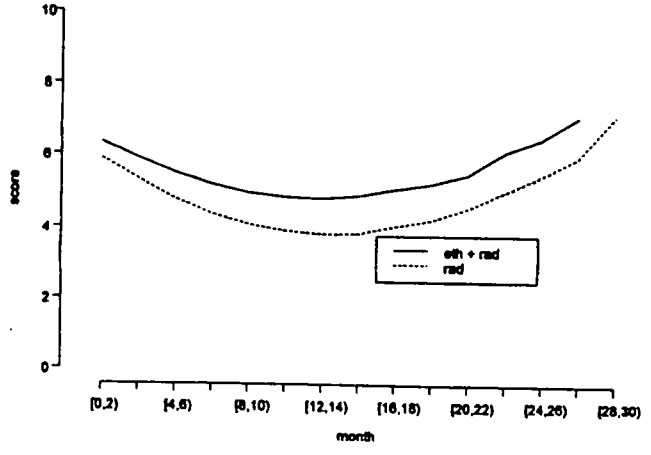
Functional Well-being



General Condition



Use of External Aids



Reviewer's Tables 8 and 9 display the results for completers. There are significant linear and quadratic time trends for the parameters "general condition" and "use of external aids" for both arms. However, for "functional well-being" the time trend remains constant in both arms.

Table 8. QOL Analyses for Completers: Ethyol + Radiation Arm

Parameter	Estimate	Standard Error *	P-value
Functional well-being			
Intercept	5.688	0.302	< 0.0001
Linear	0.005	0.056	0.925
Quadratic	0.001	0.003	0.767
General condition			
Intercept	7.138	0.294	< 0.0001
Linear	-0.226	0.057	0.0001
Quadratic	0.010	0.003	0.0002
Use of external aid			
Intercept	6.176	0.311	< 0.0001
Linear	-0.180	0.068	0.008
Quadratic	0.006	0.003	0.022

* Robust standard errors are provided.

Reviewer's Table 9. QOL Analyses for Completers: Radiation Arm

Parameter	Estimate	Standard Error *	P-value
functional well-being			
intercept	5.442	0.290	< 0.0001
linear	-0.046	0.060	0.4479
quadratic	0.003	0.002	0.2617
general condition			
intercept	6.687	0.320	< 0.0001
linear	-0.381	0.066	< 0.0001
quadratic	0.014	0.003	< 0.0001
use of external aid			
intercept	6.096	0.306	< 0.0001
linear	-0.373	0.068	< 0.0001
quadratic	0.013	0.003	< 0.0001

* Robust standard errors are provided.

These results are depicted in Reviewer's Figures 3, 4 and 5. These figures show that the time trends are different for completers and dropouts and, therefore, these groups need to be analyzed separately. Among dropouts for the parameter general condition, both arms display similar trends, where patients experienced a decline in their general condition until about the 7th month after start of therapy, then experienced an increase afterwards. Although both arms show similar trends, the patients in the Ethyol + radiation curve have noticeably higher scores than patients in the radiation alone arm. A similar phenomenon occurs among dropouts for the parameter use of external aids; however, the difference between the treatment arms is not as noticeable. Among completers the interpretation is different for functional well-being compared to the other two QOL parameters. For functional well-being patients in both arms experienced no change from their baseline score. For general condition and use of external aids, patients experienced a

decline up to a year after start of therapy, then experienced an increase afterwards. For use of external aids among completers, the patients in the Ethyol + radiation arm do not experience as dramatic a decrease in clinical benefit up to a year after start of therapy, nor do they experience as dramatic a rise as those patients in the radiation alone arm after a year. However, the increase in clinical benefit on the completers beyond the 1 year mark should be interpreted carefully due to the small number of patients returning for follow-up in the second year.

A test for difference in treatment arms can be performed, which consists of computing a t-statistic based on the difference between the estimates for trend (linear or quadratic) and a pooled standard error. Reviewer's Table 10 shows the p-values from performing such tests:

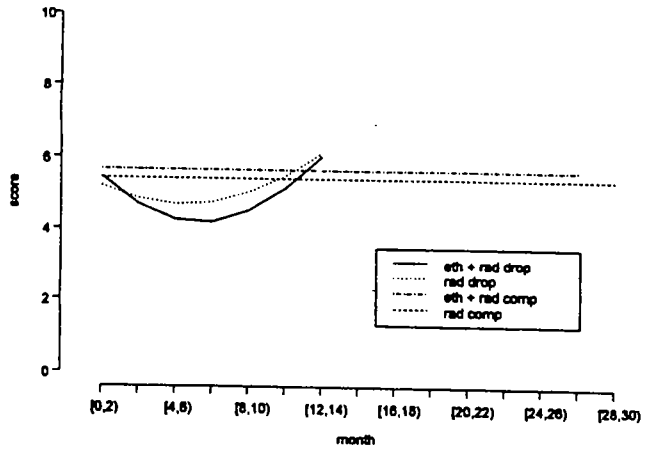
Reviewer's Table 10. Results for test of treatment difference

QOL Parameter	Linear Term p-value	Quadratic Term p-value
Functional well-being (drop)	0.0670	0.1366
Functional well-being (comp)	0.5365	0.6014
General condition (drop)	0.0083	0.0128
General condition (comp)	0.0771	0.2093
Use of external aids (drop)	0.7127	0.6313
Use of external aids (comp)	0.0446	0.0750

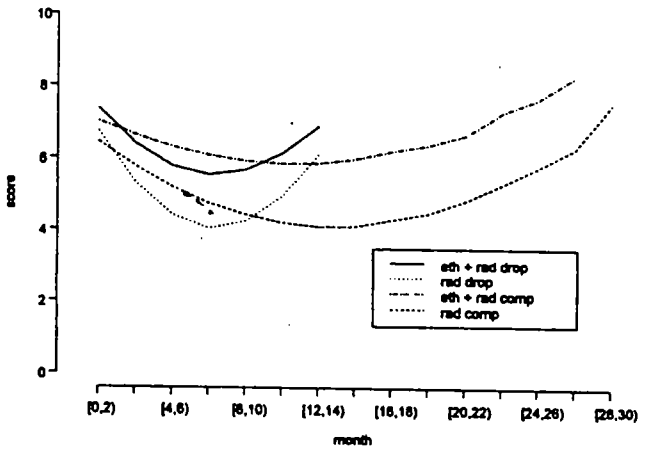
Due to the multiplicity of the tests and small sample size, the p-values provided in the above table should be interpreted with caution. Results in the above table demonstrate that the estimated time trends are similar between the two arms within dropouts and completers. The only exception would be for the parameter general condition in the dropout group. According to these results, it appears that there is not much long term clinical benefit in the Ethyol + rad group.

In addition, some QOL questions may be subjective and answers can vary depending on an individual patient's perception. For example, for the question of dryness, the responder is asked to rate from 1 to 10, the dryness of one's mouth at rest. The answer may depend on the patient's perception of dryness. Thus, the subjectivity issue in this unblinded trial setting is another factor complicating the interpretation of the QOL findings.

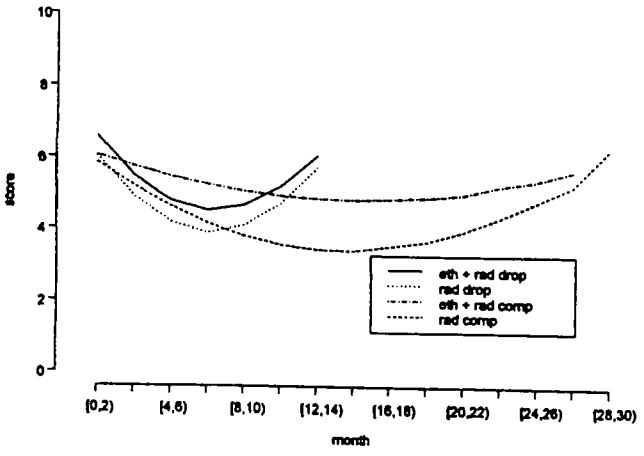
Reviewer's Figure 3. Functional well-being scores for completer and dropout groups



Reviewer's Figure 4. General condition scores for completer and dropout groups



Reviewer's Figure 5. Use of external aids scores for completer and dropout groups



Whole Unstimulated Saliva

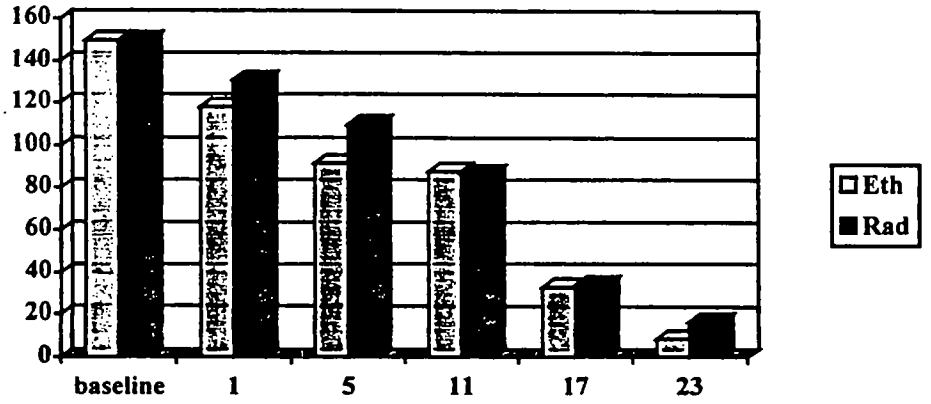
Reviewer's comment: The sponsor did not prospectively define at which time points the whole saliva measurements would be analyzed. These measurements were taken at baseline, and at 1, 5, 11, 17 and 23 months after the start of therapy. For the analysis of these multiple assessment points, it is necessary to adjust for multiple comparisons, otherwise, false positive error inflation is the consequence.

The sponsor performed a retrospective analysis on the saliva data by categorizing the measurements to be greater than or less than or equal to a cutoff value of 0.1g, which is the sponsor's definition of negligible saliva production, at the follow-up visits 1, 5, and 11. Using Fisher's exact test, they showed that a statistically significant difference ($p=0.0033$) in favor of the Ethyol + radiation arm is achieved at the time period 1 year after start of therapy. Prior to that time there was no difference between the 2 arms. Due to the retrospective selection of this cutoff, the results of this analysis should be interpreted carefully.

This reviewer analyzed the saliva measurements at all available time points using the Wilcoxon Rank Sum test. No statistically significant differences were observed at any of the time points except for the 1 year time period. The sponsor reported a p-value of 0.0419 at the 1 year time period as indication of a statistically significant difference between the treatment arms. However, due to the retrospective nature of the time point selection and test, the result should be adjusted for multiplicity and considered exploratory.

This reviewer also performed a longitudinal analysis on the saliva measurements by time. As in the case for the quality of life data, the mean value of the saliva measurements was used in the analyses for patients with measurements at multiple visits in a specific time period. This reviewer used all available data in performing the longitudinal data analysis.

Reviewer's Figure 6. Number of patients providing saliva measurements over time



In the analysis of these saliva data, the following issues are investigated: whether the dropout rates are similar in the two arms, and whether the time trends are different for the two arms for the "completers" and "dropouts". Reviewer's Figure 6 shows the frequencies of patients at

baseline and each follow-up visit. From Figure 6 it appears that both arms display similar rates of attrition. Attrition is very high in the second year of follow-up for both arms.

This reviewer used growth curve analysis to describe patterns of changes and responses over time and investigate the effect of dropouts on time trend. A GEE quadratic model was also fit to the saliva measurements. The models assume an "ignorable" missing mechanism and include both linear and quadratic terms. If the test statistics of the quadratic and/or linear terms were not significant, then those terms were dropped. The Reviewer's Tables 11, 12, and 13 summarize the results of the longitudinal analyses for the two treatment arms.

Reviewer's Table 11. Whole Unstimulated Saliva Analyses / All Patients

Treatment Arm	Parameter	Estimate	Standard Error *	P-value
Eth + Rad	intercept	2.117	0.140	< 0.0001
	linear	-0.322	0.030	< 0.0001
	quadratic	0.014	0.002	< 0.0001
Rad	intercept	2.079	0.139	< 0.0001
	linear	-0.312	0.028	< 0.0001
	quadratic	0.012	0.001	< 0.0001

* Robust standard errors are provided.

The results from Reviewer's Table 11 show that for both arms, there is a statistically significant quadratic time trend for the amount of unstimulated saliva. Figure 7 shows the estimated time profiles for each arm.

Again, a pattern mixture model (Little, 1993 and 1995) was employed to assess whether or not the missing mechanism was informative. Patients are divided into two groups: those who dropped out before the beginning of the second year of follow-up and those who remained on study through the second year of follow-up. The former group is denoted as "dropouts" and the latter "completers". This cutoff was selected because both treatment arms experienced a dropout rate of more than 50%. There are 224 (116 - Ethyol + radiation, 108 – radiation alone) dropouts and 80 (36 - Ethyol + radiation, 44 – radiation alone) completers. In Reviewer's Table 12 the results are presented for dropouts. For both arms, there are significant linear and quadratic time trends for amount of unstimulated saliva collected.

Reviewer's Table 12. Whole Unstimulated Saliva Analyses: Dropouts

Treatment Arm	Parameter	Estimate	Standard Error *	P-value
Eth + Rad	intercept	2.469	0.189	< 0.0001
	linear	-0.746	0.065	< 0.0001
	quadratic	0.054	0.005	< 0.0001
Rad	intercept	2.405	0.196	< 0.0001
	linear	-0.710	0.084	< 0.0001
	quadratic	0.050	0.006	< 0.0001

* Robust standard errors are provided.

Reviewer's Table 13. Whole Unstimulated Saliva Analyses: Completers

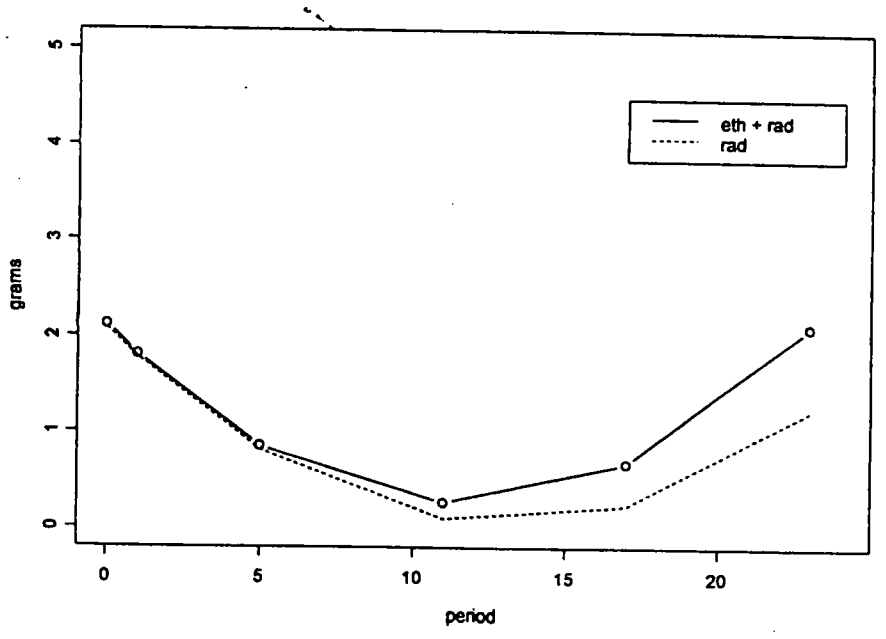
Treatment Arm	Parameter	Estimate	Standard Error *	P-value
Eth + Rad	intercept	1.808	0.223	< 0.0001
	linear	-0.242	0.039	< 0.0001
	quadratic	0.010	0.002	< 0.0001
Rad	intercept	1.885	0.228	< 0.0001
	linear	-0.242	0.035	< 0.0001
	quadratic	0.009	0.001	< 0.0001

* Robust standard errors are provided.

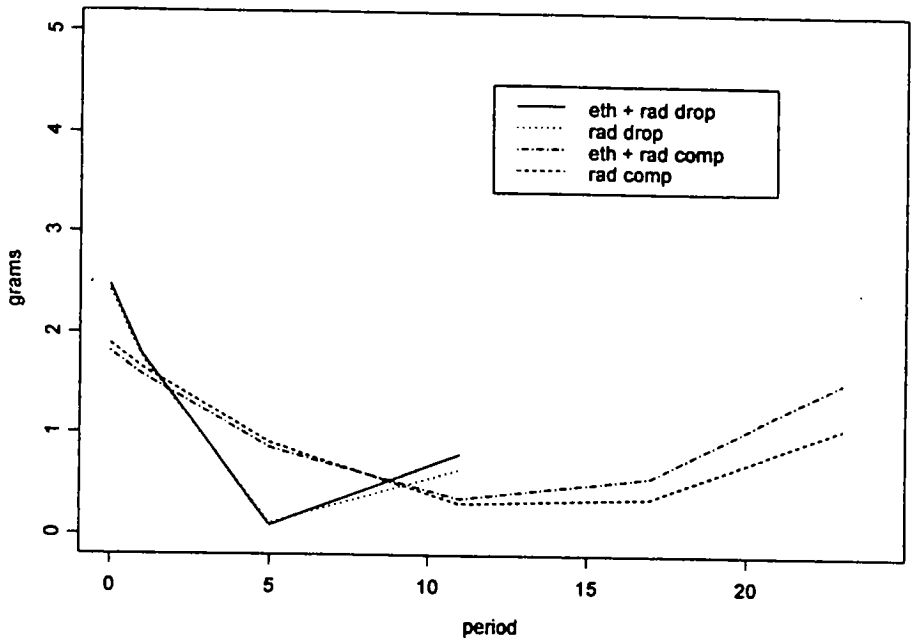
These results are depicted in Reviewer's Figure 8. The figure shows that among dropouts, there does not appear to be any difference between the treatment groups. The figure also shows among completers a slight separation between the treatment arms occurs after a year. However, this difference is not significant (p=0.9945-linear, p=0.6287-quadratic). This trend should be interpreted with caution due to the small number of patients that returned for follow-up in the second year.

According to these results, one can then conclude that there is no statistically significant difference between the treatment groups with respect to the amount of whole unstimulated saliva collected. The use of amount of whole unstimulated saliva as an endpoint may also be problematic as there are many factors that may influence the amount of saliva production and collection of an adequate specimen.

Reviewer's Figure 7. Amount of unstimulated saliva by treatment arm / All Patients



Reviewer's Figure 8. Amount of whole unstimulated saliva for completer and dropout groups



Weight loss

The sponsor's Table 19 in the appendix shows that patients in the radiation alone arm experienced greater percent of weight loss than those patients in the Ethyol + radiation arm. They claim a statistically significant difference (p=0.0437) based on the Mantel-Haenszel Chi-Square test. Since weight loss was analyzed retrospectively, the result of the analysis should be considered exploratory.

V. Overall Summary

WR-0038 was an open-label, randomized Phase III trial. The objective of this study was to determine if pre-treatment with Ethyol decreased the incidence of oral radiation toxicities without decreasing the anti-tumor efficacy of radiation. The statistical review can be summarized as follows:

1. A statistically significant difference was demonstrated between the two treatment arms with respect to incidence of greater than or equal to grade 2 acute xerostomia and late xerostomia. However, there was no difference between the two treatment arms with respect to mucositis.

2. Although no statistically significant differences were observed in the time to local regional failure, disease-free and overall survival analyses, evidence for non-inferiority is not

sufficient because of high censoring rates. More importantly, the selection of the lower limit of a one-sided 95% confidence interval of 0.7 for local regional tumor control is very liberal. To demonstrate non-inferiority, the current standard requires a lower limit of a 95% confidence interval to be at least 0.8.

3. In the analysis of the QOL parameters, there does not appear to be any statistically significant treatment difference within either dropouts or completers.

4. With respect to whole unstimulated saliva measurements, there is no statistically significant difference between treatment arms.

VI. Conclusions and Recommendations

In this supplemental NDA submission the primary efficacy endpoints are the incidence of grade 2 or higher acute xerostomia, grade 3 or higher acute mucositis, grade 2 or higher late xerostomia, and local regional tumor control rates at 12 months. The study provides evidence to support that Ethyol offers a treatment benefit for acute xerostomia and late xerostomia. However, due to an inadequate number of events for local regional control, the evidence to demonstrate non-inferiority in tumor protection for the Ethyol treatment arm is insufficient.

The analysis of the QOL parameters should be considered descriptive and exploratory because of the difficulty in interpreting the results due to the subjective nature of the questionnaire. In addition, the open-label trial design of the study can result in bias of different types. Also, adjustment for multiple comparisons is needed in order to claim a statistically significant treatment effect. Descriptively speaking, trends are in favor of the Ethyol arm for the QOL parameters general condition and use of external aids. For functional well-being it is not clear whether or not there is a trend in favor of Ethyol.

The analysis of the saliva measurements indicate that there was no difference between the treatment arms. As a result, this reviewer is unsure whether or not patients treated with Ethyol demonstrate statistically significantly better long term clinical benefit.

Clara Chu, Ph.D
Mathematical Statistician

Concur: Dr. Chen 6-14-99

Dr. Chi 6/14/99

CC: sNDA #20-221 / S-012
HFD-150/Division File
HFD-150/Ms. Pelosi, CSO
HFD-150/Dr. Chico
HFD-150/Dr. G. Williams

HFD-710/Dr. Chen
HFD-710/Dr. Chi
HFD-710/Dr. Chu
HFD-710/Chron

C:\My Documents\review\ethyol\finrev.doc
This review consists of 19 pages of text.

References

Huber, P.J. "The behavior of maximum likelihood estimates under nonstandard conditions," in *Proceedings of the fifth Berkeley Symposium in Mathematical Statistics and Probability*, Berkeley: University of California Press, 221-233, 1967.

Liang, K.Y. and Zeger, S.L. "Longitudinal data analysis using generalized linear models," *Biometrika*, **73:** 13-22, 1986.

Little, R.J.A. "Pattern-mixture models for multivariate incomplete data," *JASA*, **88:** 125-134, 1993.

Little, R.J.A. "Modeling the drop-out mechanism in repeated-measures studies," *JASA*, **90:** 1112-1121, 1995.

Westfall, P.H. and Young, S.S. "p-Value adjustments for multiple tests in multivariate binomial models," *JASA*, **84**: 780-786, 1989.

White, H. "Maximum likelihood estimation of misspecified models," *Econometrica*, **50:** 1-25, 1982.

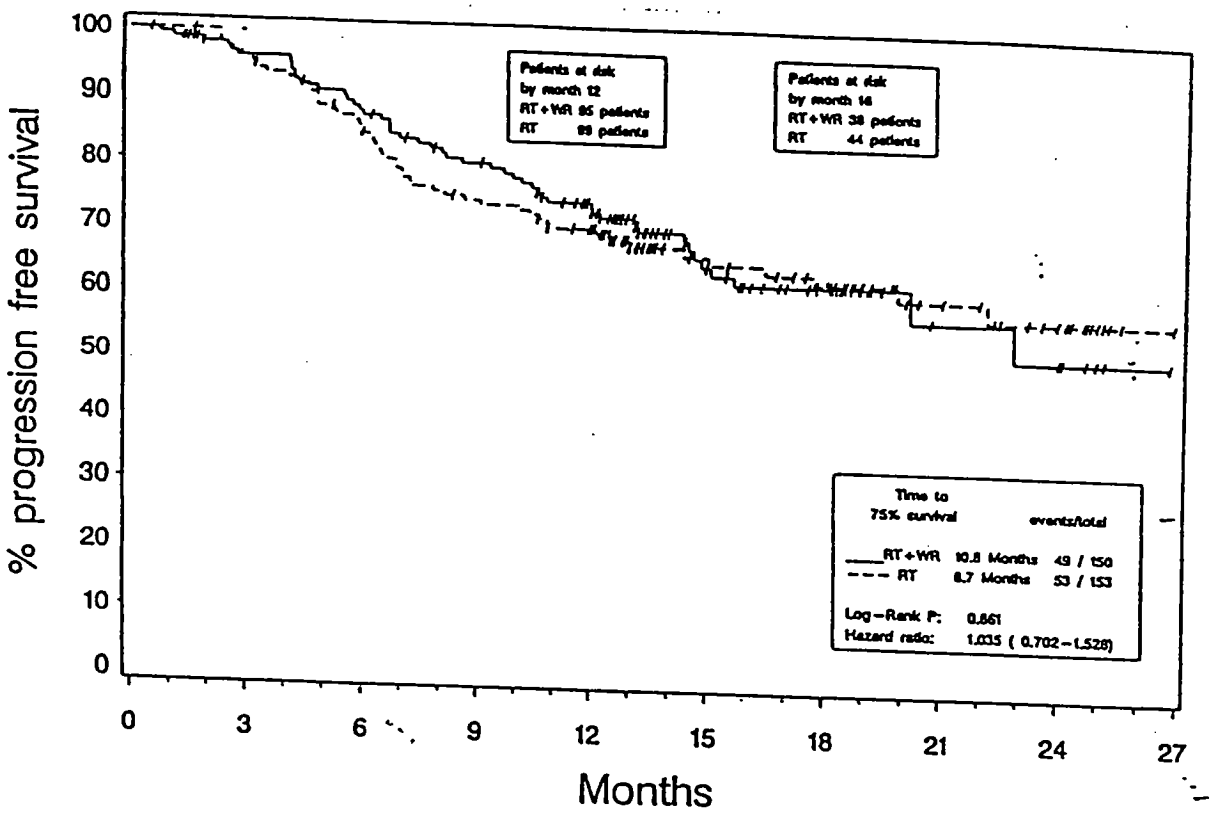


FIGURE 8: Disease-free survival of patients treated with RT ± amifostine (WR). The median follow-up was 20 months. Endpoints include progressive disease and all deaths. Time to 75% survival was 10.8 months for the amifostine + RT arm and 8.7 months for the RT alone arm. Log-rank test: 0.861; Hazard ratio and corresponding 95% confidence interval: 1.035 (0.702, 1.528).

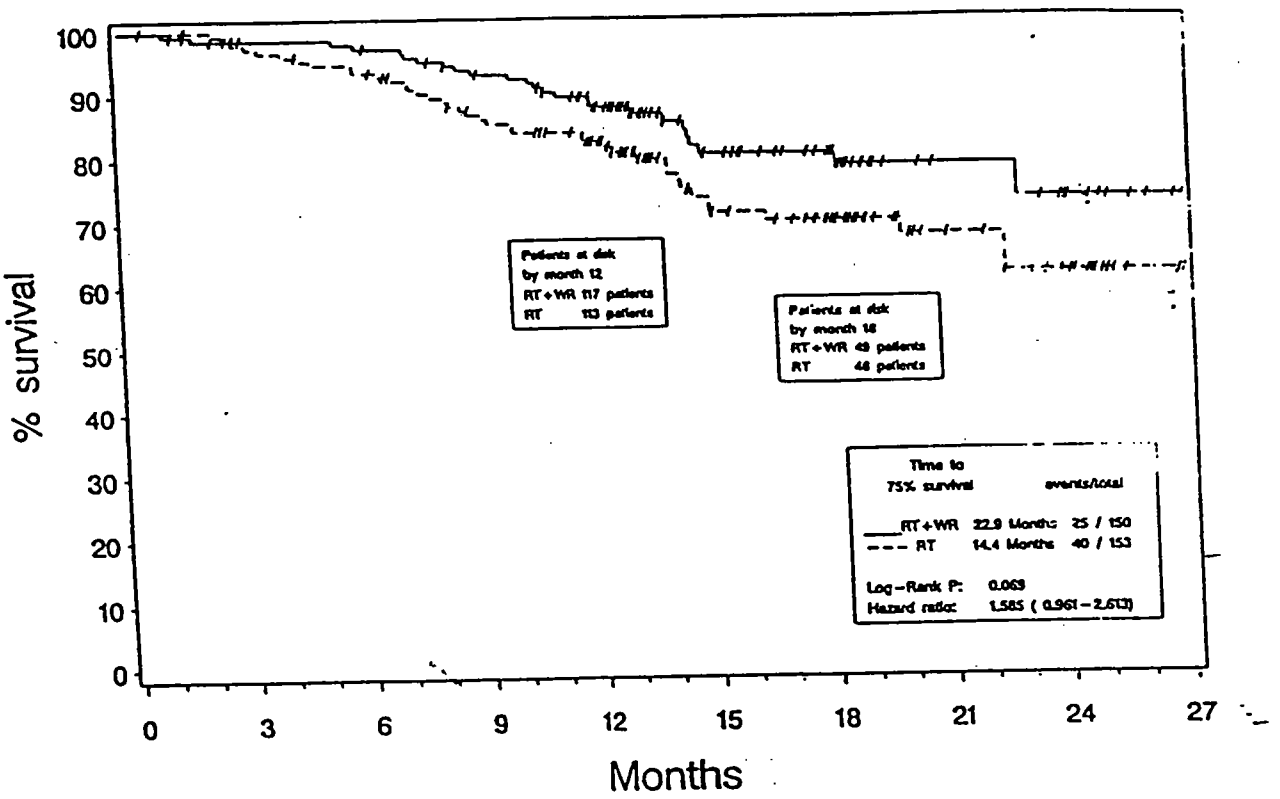


FIGURE 9: Overall survival of patients treated with RT ± amifostine (WR). Endpoints include all deaths. Time to 75% survival was 22.9 months for the amifostine + RT arm and 14.4 months for the RT alone arm. Log-rank test: 0.069; Hazard ratio and corresponding 95% confidence interval: 1.585 (0.961, 2.613).

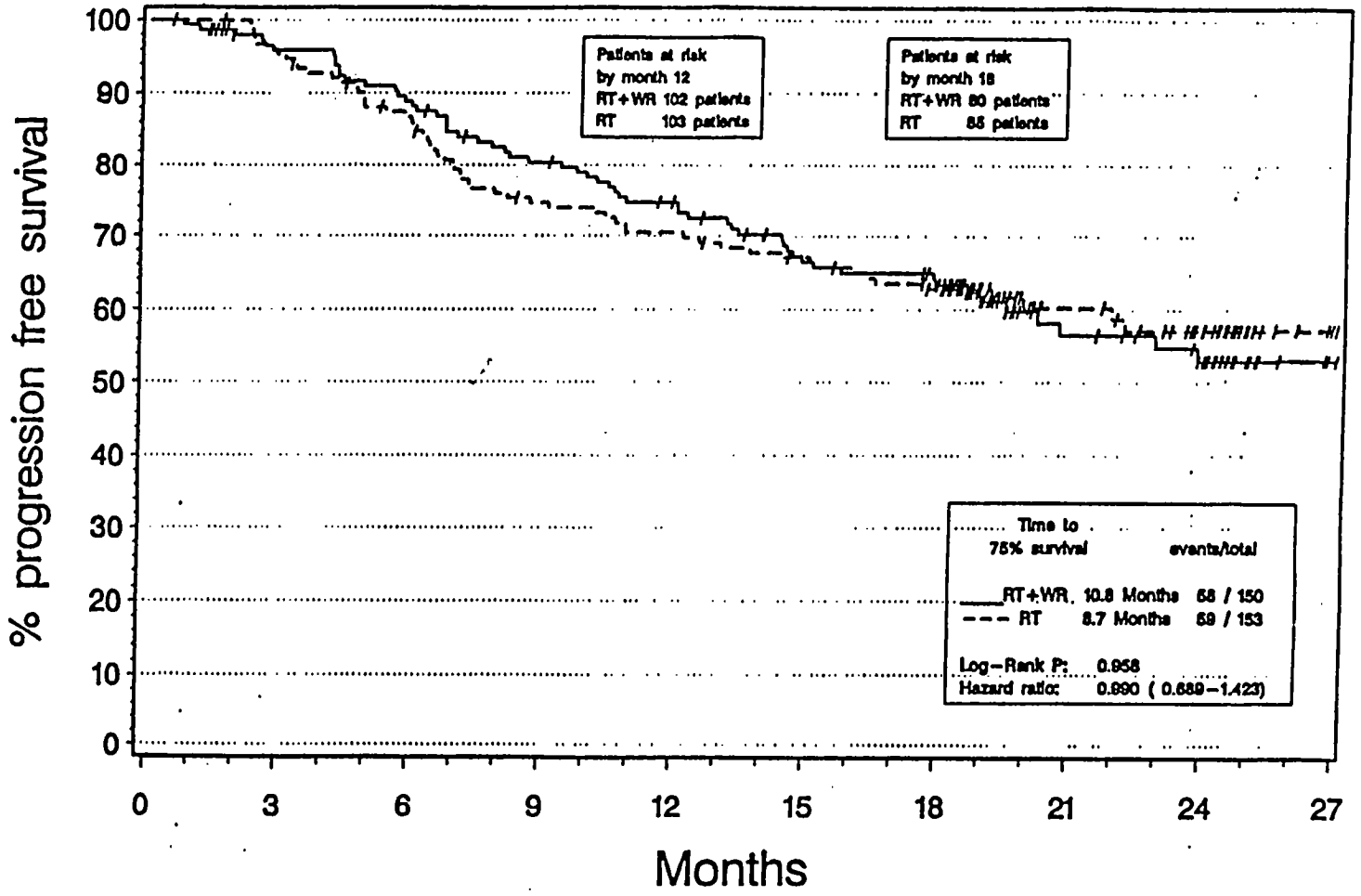


FIGURE A: 18-Month Disease-free survival curves of patients with RT ± Ethyol (WR) for head and neck cancer.

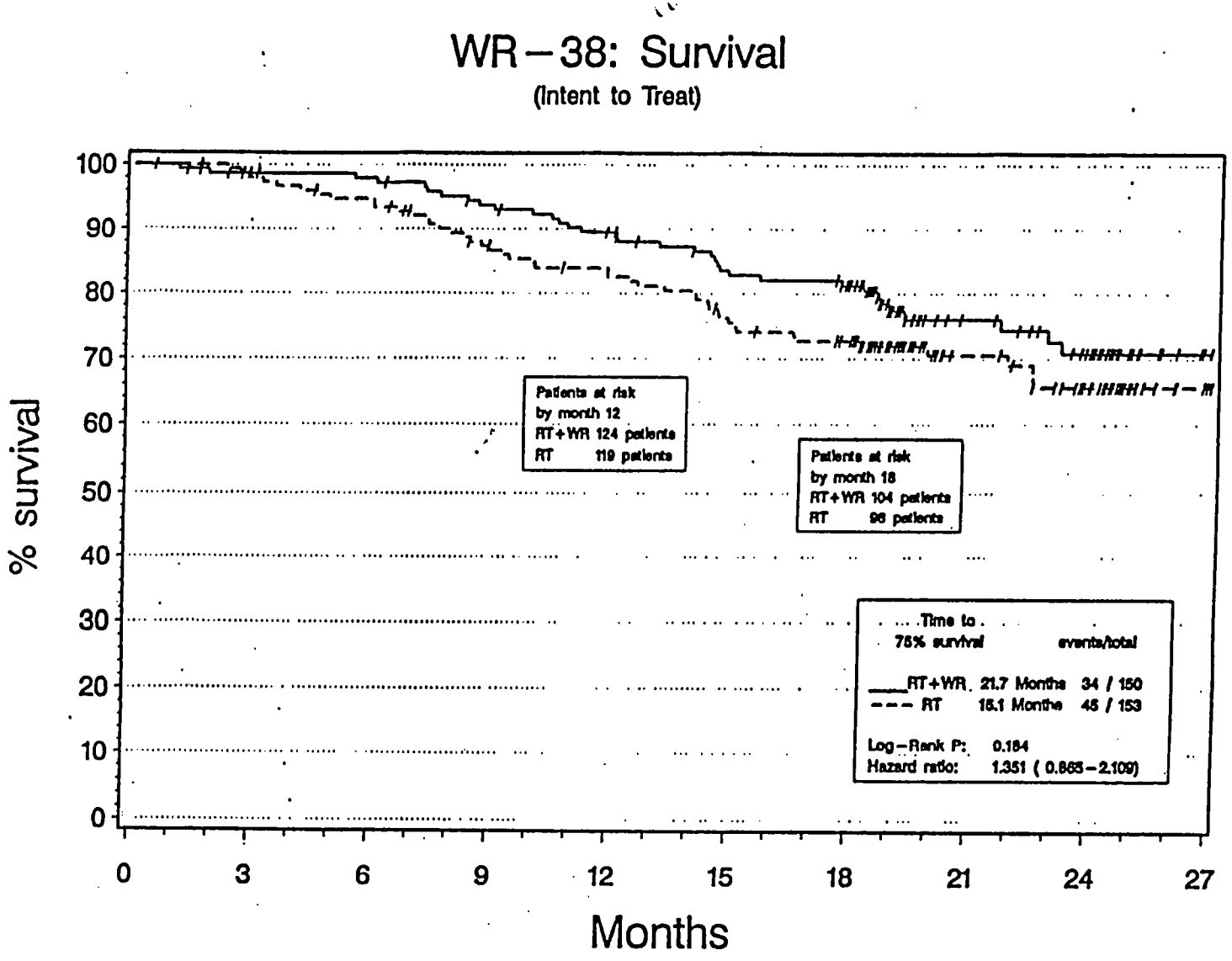


FIGURE B: 18-Month survival curves of patients with RT ± Ethyol (WR) for head and neck cancer.

TABLE 15

Comparison of PBQ Mean Scores at Each Treatment and Follow-up Visit for Patients Treated With RT ± Amifostine for Head and Neck Cancer (Intent-to-Treat Analysis)

	Amifostine + RT		RT Alone			
	Mean Score	n	Mean Score	n	Difference in Mean	p-value
Baseline	8.71	149	8.63	150	0.08	0.596
Treatment Period						
Week 1	8.31	143	8.30	147	0.01	0.967
Week 2	7.50	131	7.48	141	0.02	0.931
Week 3	7.01	129	6.51	136	0.51	0.026
Week 4	6.63	129	6.06	144	0.57	0.023
Week 5	6.24	121	5.92	141	0.32	0.214
Week 6	6.01	116	5.50	129	0.51	0.065
Week 7	5.91	90	5.30	95	0.61	0.071
End of treatment	5.90	147	5.47	151	0.43	0.098
Follow-up period						
Month 1	6.96	109	6.44	124	0.52	0.056
Month 3	6.85	103	6.55	117	0.30	0.236
Month 5	6.89	89	6.55	102	0.34	0.199
Month 7	7.39	86	6.69	89	0.70	0.009
Month 9	7.34	91	6.61	95	0.72	0.008
Month 11	7.36	83	6.66	97	0.70	0.008

TABLE 16

Comparison of Changes in PBQ Mean Scores from Baseline or First Measure at Each Treatment and Follow-up Visit for Patients Treated With RT ± Amifostine for Head and Neck Cancer (Intent-to-Treat Analysis)

	Amifostine+ RT		RT alone			
Treatment Period	Mean Change From Baseline	n	Mean Change From Baseline	n	Difference in Mean	p-value
Week 1	-0.42	143	-0.32	146	-0.10	0.430
Week 2	-1.14	131	-1.14	141	0.00	0.993
Week 3	-1.75	128	-2.08	135	0.33	0.130
Week 4	-2.10	128	-2.60	143	0.49	0.038
Week 5	-2.53	120	-2.70	140	0.17	0.518
Week 6	-2.72	115	-3.16	128	0.44	0.114
Week 7	-2.78	90	-3.22	95	0.44	0.198
Last treatment	-2.84	146	-3.15	150	0.31	0.242
Follow-up Period						
Month 1	-1.81	109	-2.17	123	0.36	0.189
Month 3	-2.03	102	-2.07	117	0.03	0.907
Month 5	-2.05	89	-2.18	102	0.13	0.657
Month 7	-1.59	85	-2.06	89	0.47	0.129
Month 9	-1.58	91	-2.17	94	0.60	0.033
Month 11	-1.62	83	-1.99	96	0.37	0.180

Figure 2. Graph of Laird–Ware model of PBQ scores for completers

Graph shows PBQ scores through time for Amifostine (WR+RT) and control (PT) patients. A treatment effect is shown in completers.

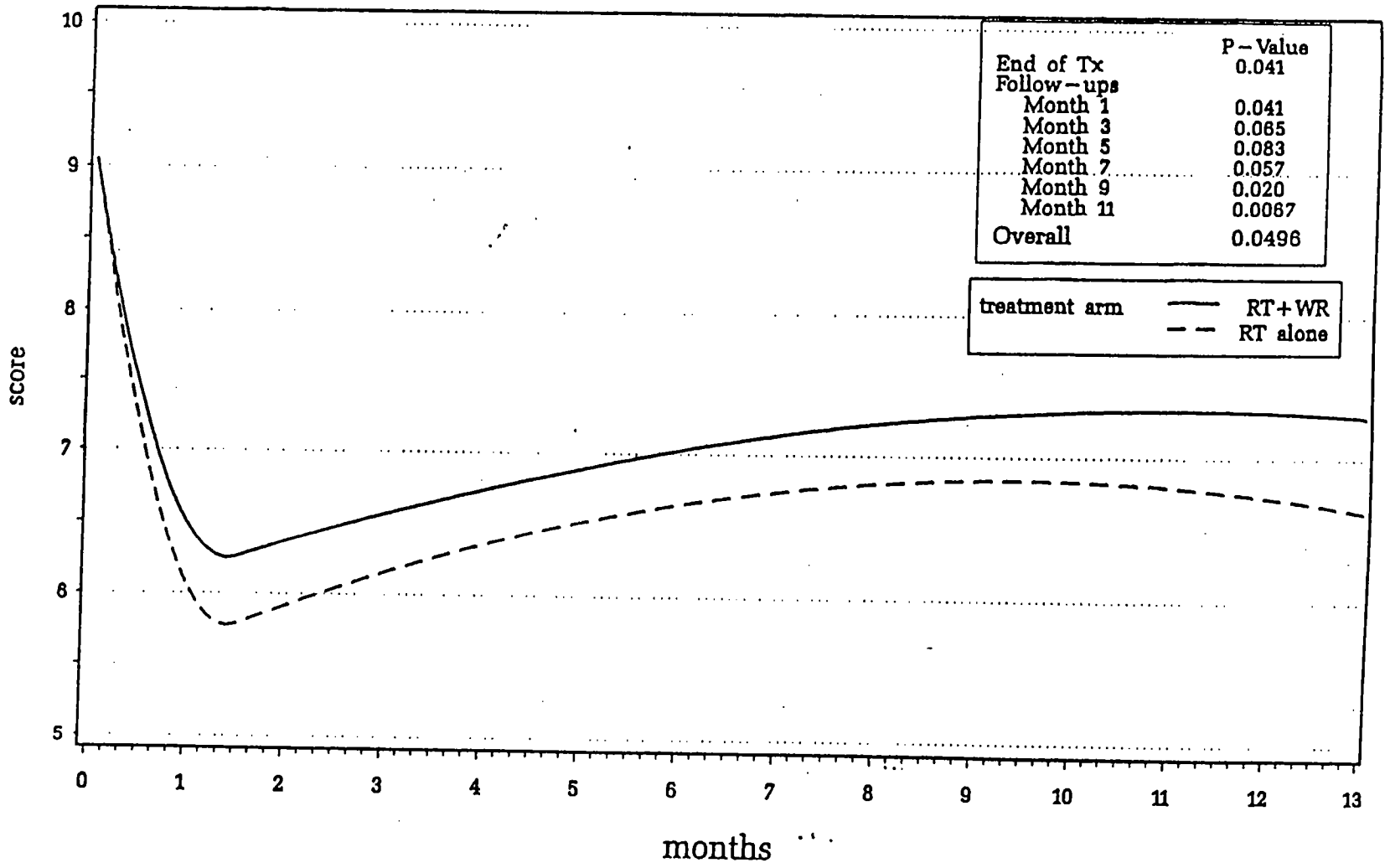


Figure 3. Graph of Laird – Ware model of PBQ scores for non – completers

Graph shows PBQ scores through time for Amifostine (WR+RT) and control (PT) patients.
A treatment effect is not shown in non – completers.

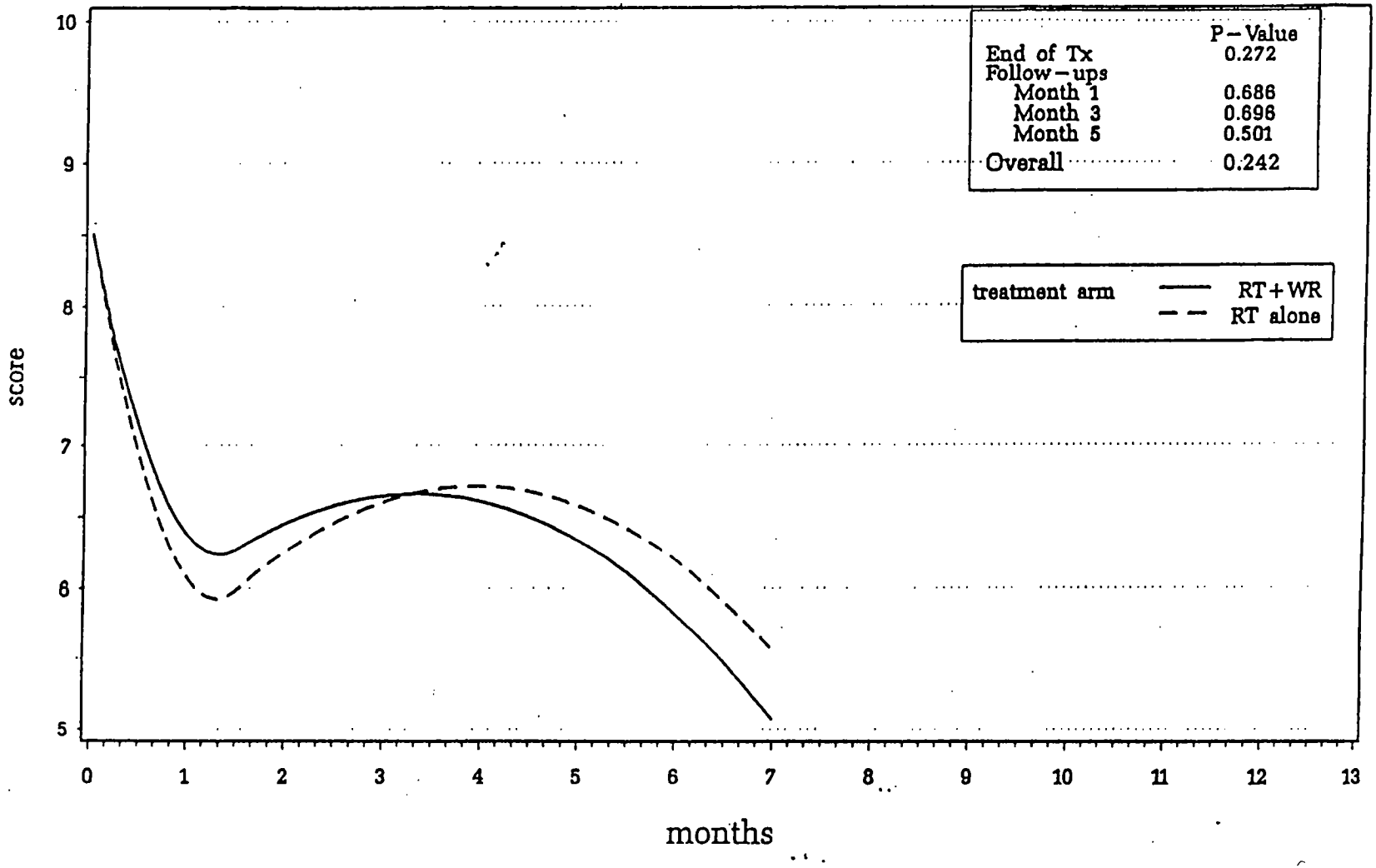


TABLE 19

Percent Weight Loss From Baseline in Patients Treated With RT ± Amifostine for Head and Neck Cancer

Percent Weight Loss	Amifostine + RT (N=150)		RT Alone (N=153)		p-value
None	35	(24%)	21	(14%)	
<5%	42	(29%)	47	(31%)	
5%-10%	44	(30%)	53	(35%)	
>10%	24	(17%)	32	(21%)	0.0437[a]
Total ≥5%	68	(47%)	85	(56%)	0.1641[b]

[a] Based on Mantel-Haenszel Chi-square test
[b] Based on Fisher exact test.

Table A1. QOL Analysis of Minimum Scores: All patients

Treatment Group	Parameter	Estimate	Standard Error	P-value
Ethyol + Radiation	functional well-being			
	intercept	5.524	0.182	< 0.0001
	linear	-0.099	0.046	0.030
	quadratic	0.006	0.003	0.0160
Ethyol + Radiation	general condition			
	intercept	7.345	0.163	< 0.0001
	linear	-0.282	0.042	< 0.0001
	quadratic	0.013	0.002	< 0.0001
Ethyol + Radiation	use of external aid			
	intercept	6.502	0.185	< 0.0001
	linear	-0.267	0.048	< 0.0001
	quadratic	0.011	0.002	< 0.0001
Radiation	functional well-being			
	intercept	5.217	0.187	< 0.0001
	linear	-0.026	0.040	0.522
	quadratic	0.003	0.002	0.152
Radiation	general condition			
	intercept	6.770	0.172	< 0.0001
	linear	-0.405	0.041	< 0.0001
	quadratic	0.016	0.002	< 0.0001
Radiation	use of external aid			
	intercept	6.084	0.179	< 0.0001
	linear	-0.338	0.042	< 0.0001
	quadratic	0.013	0.002	< 0.0001

Table A2. QOL Analysis of Minimum Scores: Dropouts

Treatment Group	Parameter	Estimate	Standard Error	P-value
Ethyol + Radiation	functional well-being			
	intercept	5.736	0.239	< 0.0001
	linear	-0.513	0.102	< 0.0001
	quadratic	0.042	0.008	< 0.0001
Ethyol + Radiation	general condition			
	intercept	7.684	0.190	< 0.0001
	linear	-0.606	0.085	< 0.0001
	quadratic	0.042	0.007	< 0.0001
Ethyol + Radiation	use of external aid			
	intercept	6.984	0.230	< 0.0001
	linear	-0.701	0.101	< 0.0001
	quadratic	0.050	0.008	< 0.0001
Radiation	functional well-being			
	intercept	5.340	0.257	< 0.0001
	linear	-0.262	0.097	0.007
	quadratic	0.026	0.007	0.0003
Radiation	general condition			
	intercept	7.260	0.199	< 0.0001
	linear	-0.922	0.085	< 0.0001
	quadratic	0.066	0.007	< 0.0001
Radiation	use of external aid			
	intercept	6.461	0.230	< 0.0001
	linear	-0.749	0.085	< 0.0001
	quadratic	0.055	0.007	< 0.0001

Table A3. QOL Analysis of Minimum Scores: Completers

Treatment Group	Parameter	Estimate	Standard Error	P-value
Ethyol + Radiation	functional well-being			
	intercept	5.662	0.301	< 0.0001
	linear	0.010	0.056	0.863
	quadratic	0.001	0.003	0.810
Ethyol + Radiation	general condition			
	intercept	7.112	0.297	< 0.0001
	linear	-0.222	0.057	0.0001
	quadratic	0.009	0.003	0.0003
Ethyol + Radiation	use of external aid			
	intercept	6.140	0.311	< 0.0001
	linear	-0.175	0.068	0.010
	quadratic	0.006	0.003	0.026
Radiation	functional well-being			
	intercept	5.382	0.293	< 0.0001
	linear	-0.036	0.060	0.551
	quadratic	0.002	0.002	0.323
Radiation	general condition			
	intercept	6.629	0.322	< 0.0001
	linear	-0.372	0.067	< 0.0001
	quadratic	0.014	0.003	< 0.0001
Radiation	use of external aid			
	intercept	6.041	0.310	< 0.0001
	linear	-0.364	0.068	< 0.0001
	quadratic	0.013	0.003	< 0.0001

Table B1. QOL Analysis of Maximum Scores: All Patients

Treatment Group	Parameter	Estimate	Standard Error	P-value
Ethyol + Radiation	functional well-being			
	intercept	5.598	0.181	< 0.0001
	linear	-0.113	0.045	0.013
	quadratic	0.007	0.003	0.009
Ethyol + Radiation	general condition			
	intercept	7.399	0.162	< 0.0001
	linear	-0.289	0.042	< 0.0001
	quadratic	0.013	0.002	< 0.0001
Ethyol + Radiation	use of external aid			
	intercept	6.574	0.185	< 0.0001
	linear	-0.280	0.048	< 0.0001
	quadratic	0.012	0.002	< 0.0001
Radiation	functional well-being			
	intercept	5.301	0.186	< 0.0001
	linear	-0.039	0.040	0.328
	quadratic	0.003	0.002	0.099
Radiation	general condition			
	intercept	6.854	0.170	< 0.0001
	linear	-0.420	0.041	< 0.0001
	quadratic	0.017	0.002	< 0.0001
Radiation	use of external aid			
	intercept	6.179	0.176	< 0.0001
	linear	-0.354	0.042	< 0.0001
	quadratic	0.014	0.002	< 0.0001

Table B2. QOL Analysis of Maximum Scores: Dropouts

Treatment Group	Parameter	Estimate	Standard Error	P-value
Ethyol + Radiation	functional well-being			
	intercept	5.829	0.235	< 0.0001
	linear	-0.537	0.100	< 0.0001
	quadratic	0.043	0.008	< 0.0001
Ethyol + Radiation	general condition			
	intercept	7.743	0.190	< 0.0001
	linear	-0.618	0.084	< 0.0001
	quadratic	0.043	0.007	< 0.0001
Ethyol + Radiation	use of external aid			
	intercept	7.058	0.230	< 0.0001
	linear	-0.720	0.100	< 0.0001
	quadratic	0.051	0.008	< 0.0001
Radiation	functional well-being			
	intercept	5.407	0.257	< 0.0001
	linear	-0.273	0.097	0.0051
	quadratic	0.027	0.007	0.0003
Radiation	general condition			
	intercept	7.331	0.196	< 0.0001
	linear	-0.939	0.086	< 0.0001
	quadratic	0.066	0.007	< 0.0001
Radiation	use of external aid			
	intercept	6.555	0.227	< 0.0001
	linear	-0.769	0.086	< 0.0001
	quadratic	0.056	0.007	< 0.0001

Table B3. QOL Analysis of Maximum Scores: Completers

Treatment Group	Parameter	Estimate	Standard Error	P-value
Ethyol + Radiation	functional well-being			
	intercept	5.714	0.303	< 0.0001
	linear	0.001	0.056	0.988
	quadratic	0.001	0.003	0.723
Ethyol + Radiation	general condition			
	intercept	7.164	0.292	< 0.0001
	linear	-0.230	0.057	0.0001
	quadratic	0.010	0.003	0.0002
Ethyol + Radiation	use of external aid			
	intercept	6.212	0.313	< 0.0001
	linear	-0.186	0.067	0.006
	quadratic	0.007	0.003	0.018
Radiation	functional well-being			
	intercept	5.502	0.287	< 0.0001
	linear	-0.055	0.060	0.357
	quadratic	0.003	0.002	0.209
Radiation	general condition			
	intercept	6.744	0.320	< 0.0001
	linear	-0.390	0.066	< 0.0001
	quadratic	0.014	0.003	< 0.0001
Radiation	use of external aid			
	intercept	6.151	0.304	< 0.0001
	linear	-0.382	0.068	< 0.0001
	quadratic	0.014	0.003	< 0.0001