CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER: NDA 20449/S11

STATISTICAL REVIEW(S)

Statistical Review and Evaluation

NDA #:	20-449
Sponsor:	Rhône-Poulenc Rorer
Name of Drug:	Taxotere (docetaxel) for Injection Concentrate
Indication:	Treatment of NSCLC in patients who've been previously treated w/ platinum- based chemotherapy
Documents Reviewed:	Volumes 1-6, dated 12/23/98; volumes 1-5, dated 6/23/99
Medical Reviewer:	Donna Griebel, M.D.

This review consists of eleven sections. The first section provides some brief background information, and lists the major statistical issues. The second, third and fourth sections provide a description of the pivotal study TAX 320, and present the primary and secondary efficacy results, except the QOL analyses. The fifth, sixth, and seventh sections provide a description of the pivotal study TAX 317, and present the primary and secondary efficacy results, except the QOL analyses. In section eight the QOL analyses and results for both pivotal studies are presented. Section nine contains the results of exploratory survival analyses. In section ten, 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

This supplemental NDA is comprised of 2 pivotal Phase III trials: TAX 320 and TAX 317. In addition, 4 supportive single agent Phase II trials (TAX 270, TAX 271, TAX 297, SI002A) at a dose of 100mg/m^2 and 2 additional studies with one at a dose of 60 mg/m^2 (TAX241) and the other at a dose of 75 mg/m^2 (CHI202) were also submitted. This statistical review will only focus on studies TAX 320 and TAX 317. Some major statistical issues have been identified by this reviewer:

1. An inappropriate statistical method is used to compare the 1-year survival estimates.
2. In study TAX 317 the primary efficacy results are inappropriately based on a subgroup analysis.
3. Inappropriate analyses for QOL data were performed.

II. Description of TAX 320

Objective

The objective of this study was to compare survival in patients with non-small cell lung cancer (NSCLC) previously treated with platinum-containing chemotherapy receiving 100 or 75 mg/m^2 docetaxel vs. either vinorelbine 30mg/m^2 per week or ifosfamide 6 mg/m^2.

Design

TAX320 is a multi-center (U.S. sites), open-label, randomized parallel group Phase III study comparing three arms:

1. Docetaxel (100 mg/m^2 i.v. infusion on Day 1, repeated every 21 days)
2. Docetaxel (75 mg/m^2 i.v. infusion on Day 1, repeated every 21 days)
3. Either vinorelbine (30 mg/m^2 i.v. infusion on Days 1, 8, and 15 of each 3 week cycle) or ifosfamide (2 mg/m^2 on Days 1-3 of every 3 week cycle)

for treatment of patients who have previously been treated with a platinum-based regimen. Prior to central randomization, patients were stratified on the following factors:

1. Best response to prior platinum therapy (progression vs. NC, PR, CR)
2. ECOG status (0-1 vs. 2).

Patients were treated for 6 cycles or until evidence of progressive disease or unacceptable side-effects. Patients were assessed at every cycle, within 30 days after the last treatment, and followed up every 2 months after the last treatment cycle until death.

Patient Population (Protocol)

The protocol specified a sample size of 360 eligible and evaluable patients with 120 randomized into each treatment arm. The sample size is based on the following assumptions:

1. the median survival times for either docetaxel dose and the control group are 7.5 and 5 months respectively
2. the patient accrual time is 9 months
3. the follow-up time after the last patient accrued is 12 months.

Given the above assumptions, the sample size of 120 patients in each arm will result in an alpha level of 0.05 (1-sided) and power of 80%. The sample size is adjusted for an overall 10% dropout. Although a one-sided error rate was the basis of the determination of the sample size, all statistical tests are based on a two-sided error rate of 5%.

Efficacy Endpoints

The primary efficacy endpoint is survival time, which is measured from the randomization date to the date of death from any cause or the date of last contact if there was no documentation of death.

The secondary efficacy endpoints include percentage of patients with objective response rate, duration of tumor response, time to progression, and quality of life measurements. Time to progression is defined from the randomization date to the date of first determination of progressive disease or to the date of the last assessment prior to antitumor therapy (including radiotherapy). Quality of life measurements are obtained from the Lung Cancer Symptom Scale (LCSS). The endpoints defined from the LCSS are changes in LCSS scores from baseline scores rated on a 100 mm visual analog scale, and changes in ECOG performance status, body weight, and analgesic use.

Statistical Methods (Protocol)

1. The logrank and Kaplan Meier methods were used to analyze survival and time to progression. The median and Kaplan-Meier estimates were given with their 95% confidence intervals. Comparison between treatment groups was performed using the logrank test stratifying on ECOG performance status and response to prior platinum therapy. The Bonferroni method was used to adjust for multiple comparisons.

2. Chi-square tests were used to compare response rates (replaced by Fisher's exact test if the expected frequency in one cell of the table is < 5). The 95% confidence interval for proportions was calculated following the exact method.

3. For the QOL endpoints, the sponsor proposed a missing data imputation strategy. Then, the QOL endpoints were analyzed using a longitudinal model and the GEE approach to test

whether or not the assumption of randomly missing information holds. If the assumption does hold, then an analysis of covariance will be employed for the endpoint analysis using the last value carried forward method (LVCF) on the total scores and the subscale scores, while using a set of prospectively defined covariates.

Study Patients

The study population consists of 373 patients. These patients constitute the intent to treat (ITT) population, which is all patients randomized. The ITT population for time to progression and best overall response (BOR) includes randomized patients diagnosed with NSCLC since progression and BOR due to NSCLC could not be determined in patients not diagnosed with NSCLC. For time to progression and BOR, there are 370 patients. Three patients, 1 in each group, were excluded because they did not have NSCLC. The sponsor also defined an evaluable patient subgroup consisting of 317 patients, who were evaluable for response.

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

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

Survival Analysis

Overall survival was defined from the randomization date to the date of death from any cause. Patients were censored at the date of last contact if there was no documentation of death or at the cutoff date if death occurred after this date. Survival was analyzed using the Kaplan Meier method and the logrank test. The test did not reveal any statistically significant difference between each of the treatment arms and the control (p=0.93 - docetaxel 100mg/m^2 vs. control, p=0.14 – docetaxel 75mg/m^2 vs. control). The results are summarized in Reviewer's Table 1 below and in the Sponsor's Figure 4.01AF:

Reviewer's Table 1. Sponsor's Survival Analysis (ITT)

	Docetaxel 100mg/m^2	Docetaxel 75mg/m^2	Vinorelbine/ Ifosfamide
Patients randomized	125	125	123
Number of deaths	104 (83.2%)	97 (77.6%)	110 (89.4%)
Number censored	21 (16.8%)	28 (22.4%)	13 (10.6%)
Median survival (months)	5.5	5.7	5.6
95% CI (months)	4.6, 6.6	5.1, 7.9	4.3, 7.9
1 yr. Survival K-M estimate (%)	21	32	19
95% CI (%)	14, 28	23, 40	12, 26
Logrank test vs. comparator	p=0.93	p=0.14	
Logrank test (D/100=D/75=V/I)	Overall p=0.26		

The sponsor also reported that *of the patients who were not lost to follow-up*, the proportion of patients alive at 1 year was significantly different favoring the D/75 treatment group (32 alive, 83 deceased) compared to the V/I treatment group (20 alive, 98 deceased) [p=0.046, chi-square test].

Reviewer's Comment:

1. Retrospectively choosing a particular time point to analyze the survival data is not appropriate. Results based on such an analysis should be considered exploratory.
2. It is not appropriate to do a proportional test for patients alive at 1 year. This analysis ignores the degree and pattern of censoring which may lead to bias in the comparison of two proportions. The logrank test should be performed in order to test for a difference between treatment groups by censoring individuals who survive beyond 1 year. The results of this reviewer's analysis are given in Reviewer's Table 2.

Reviewer's Table 2. Reviewer's Survival Analysis to Test for Treatment Difference at 1 Year Survival (ITT)

	Docetaxel 100mg/m^2	Docetaxel 75mg/m^2	Vinorelbine/ Ifosfamide
Patients randomized	125	125	123
Number of deaths	96 (76.8%)	83 (66.4%)	98 (79.7%)
Number censored	29 (23.2%)	42 (33.6%)	25 (20.3%)
Logrank test vs. comparator	0.79	0.20	

The results yield p-values that are not significant in comparing either docetaxel group to the control group (p=0.79 - D/100 vs. V/I, p=0.20 – D/75 vs. V/I).

The sponsor also performed the survival analysis while censoring the patients at the time of subsequent chemotherapy. The reason for performing this analysis was because of the "large proportion of patients having received post-study chemotherapy and the difference in drug exposure of post-study chemotherapy was substantially different between treatment groups." The results are summarized in Reviewer's Table 3:

Reviewer's Table 3. Sponsor's Survival Analysis (ITT) – Censoring at Subsequent Chemotherapy

	Docetaxel 100mg/m^2	Docetaxel 75mg/m^2	Vinorelbine/ Ifosfamide
Patients randomized	125	125	123
Number of deaths	68 (54.4%)	64 (51.2%)	72 (58.5%)
Number censored	57 (45.6%)	61 (48.8%)	51 (41.5%)
Median survival (months)	6.6	5.8	5.4
95% CI (months)	5.0, 7.9	5.2, 8.0	4.2, 7.9
1 yr. Survival K-M estimate (%)	32	32	10
95% CI (%)	22, 43	20, 44	1, 18
Logrank test vs. comparator	p=0.25	p=0.12	
Logrank test (D/100=D/75=V/I)	Overall p=0.22		

Because the survival estimates for the D/100 and the D/75 groups were similar, the two dose groups were combined. The results are summarized in Reviewer's Table 4:

Reviewer's Table 4. Sponsor's Survival Analysis (ITT) – Combined Docetaxel Doses with Censoring at Subsequent Chemotherapy

	Docetaxel	Vinorelbine/ Ifosfamide
Patients randomized	250	123
Number of deaths	132 (52.8%)	72 (58.5%)
Number censored	118 (47.2%)	51 (41.5%)
Median survival (months)	6.5	5.4
95% CI (months)	5.4, 7.6	4.2, 7.9
1 yr. Survival K-M estimate (%)	32	10
95% CI (%)	24, 40	1, 18
Logrank test vs. comparator	p=0.08	

The sponsor also reported that of the patients who were not censored at the 1 year timepoint, the proportion of patients alive at 1 year is significantly different in favor of the docetaxel group (26 alive, 125 dead) vs. the V/I group (4 alive, 70 dead). This was done using the Chi-square test (p=0.012).

Reviewer's Comment:

1. Results of the survival analyses with censoring patients at subsequent therapy and combining doses should be considered supportive and exploratory.
2. As discussed in the reviewer's comment regarding the proportion analysis for 1 year survival, the proportional test for 1 year survival is not appropriate. This reviewer's results, based on the logrank test for 1 year survival, are summarized in Reviewer's Table 5.

Reviewer's Table 5. Reviewer's Survival Analysis to Test for Treatment Difference at 1 Year Survival (ITT) – Combined Docetaxel Doses with Censoring at Subsequent Chemotherapy

	Docetaxel	Vinorelbine/ Ifosfamide
Patients randomized	250	123
Number of deaths	125 (50%)	70 (56.9%)
Number censored	125 (50%)	53 (43.1%)
Logrank test vs. comparator	p=0.095	

The results yield a p-value that is not significant in comparing the combined docetaxel group to the control group (p=0.095 – docetaxel vs. V/I).

IV. Secondary Efficacy Results

Response Rate

Overall response rate is the sum of all confirmed complete and partial (CR + PR) responses in each treatment group. The response rate for the D/100 treatment group was found to be statistically significantly greater than that of the V/I treatment group in the ITT population (10.5% vs. 0.8% using Fisher's Exact test, p=0.001) and for the evaluable population (11.9% vs.

1%, p=0.001). For the D/75 treatment group, the response rate was also statistically significant in both the ITT population (6.5% vs. 0.08%, p=0.036) and the evaluable population (7.5% vs. 1%, p=0.036).

Reviewer's Comment: Since there are multiple comparisons in the analysis of response rate, the p-values need to be appropriately adjusted. For example, if using the Bonferroni adjustment, the treatment difference between the D/75 arm and V/I arm will no longer be statistically significant (p=0.072).

The MO reclassified the response rates for some patients (see MO review). These changes gave the following results:

Reviewer's Table 6. MO's RR Analysis (ITT)

	Docetaxel 100mg/m^2	Docetaxel 75mg/m^2	Vinorelbine/ Ifosfamide
Patients randomized	124	124	122
Overall RR	11 (8.9%)	6 (4.8%)	1 (0.8%)
Fishers Exact test vs. comparator	p=0.005	p=0.12	

Based on Reviewer's Table 6, there remains a statistically significant treatment difference for the D/100 group compared to the control group, but not for the D/75 group vs. the control group.

Time to Progression

Time to progression (TTP) was analyzed for each treatment group using the Kaplan Meier method. TTP was defined from the randomization date to the date of documented progressive disease. Patients were censored at the time of the last tumor assessment before further anticancer therapy, excluding radiotherapy. The results are summarized in Reviewer's Table 7:

The logrank test result of p=0.044 shows that there is a statistically significant difference when comparing D/100 and V/I. After pooling the two docetaxel doses, the logrank test gives p=0.046.

Reviewer's Table 7. Sponsor's Time to Progression Analysis (ITT) – Censored at subsequent chemotherapy

	Docetaxel 100mg/m^2	Docetaxel 75mg/m^2	Vinorelbine/ Ifosfamide
Patients randomized	124	124	122
Number of PD	115 (92.7%)	117 (94.4%)	114 (93.4%)
Number censored	9 (7.3%)	7 (5.6%)	8 (6.6%)
Median time to PD (weeks)	8.4	8.5	7.9
95% CI (weeks)	6.7, 11.0	6.7, 11.0	6.9, 11.0
26 week K-M estimate (%)	19	17	8
95% CI (%)	12, 26	10, 24	3, 13
Logrank test vs. comparator	p=0.044	p=0.093	

Reviewer's Comment: After adjusting for multiple comparisons, the treatment difference of D/100 vs. V/I may not be statistically significant (e.g., if using Bonferroni adjustment, p=0.088).

The MO did not believe that patients should be censored at the time of further anticancer therapy, but instead, they should have been counted as events at this time. In addition, there were errors in the assignment of PD dates for many patients. As a result, changes were made reflecting the MO's and the sponsor's comments from the 11/5/99 communication between the FDA and the sponsor. An exploratory analysis was performed by this reviewer to determine the effect that these changes had on the TTP analysis. The results are shown in Reviewer's Table 8:

Reviewer's Table 8. FDA Reviewer's Time to Progression Analysis (ITT)

	Docetaxel 100mg/m^2	Docetaxel 75mg/m^2	Vinorelbine/ Ifosfamide
Patients randomized	124	124	122
Number of PD	118 (95.2%)	122 (98.4%)	119 (97.5%)
Number censored	6 (4.8%)	2 (1.6%)	3 (2.5%)
Median time to PD (weeks)	8.4	8.3	7.6
95% CI (weeks)	7.0, 10.1	6.7, 11.7	6.7, 10.1
26 week K-M estimate (%)	16	15	7
95% CI (%)	10, 23	9, 22	2, 11
Logrank test vs. comparator	p=0.064	p=0.074	

The updated TTP now shows that neither docetaxel arm is statistically significant.

V. Description of TAX 317

Objective

The objective of this study was to compare survival in patients with non-small cell lung cancer (NSCLC) previously treated with platinum-containing chemotherapy receiving either docetaxel or best supportive care.

Design

TAX 317 is a multi-center (U.S., Canadian, European sites), open-label, randomized parallel group Phase III study comparing two arms:

1. Docetaxel (100 mg/m^2 - later reduced to 75 mg/m^2 i.v. infusion on Day 1, repeated every 21 days)
2. Best supportive care; no systemic chemotherapy or any systemic anticancer therapy allowed for treatment of patients who have previously been treated with platinum-containing chemotherapy. Prior to central randomization, patients were stratified on the following factors:
1. Best response to prior platinum therapy (progression vs. NC, PR, CR)
2. ECOG status (0-1 vs. 2).

Patients were treated until evidence of progressive disease or unacceptable side-effects. Patients were assessed at every cycle, within 30 days after the last treatment, and followed up every 2 months after the last treatment cycle until death.

Patient Population (Protocol)

The protocol specified a sample size of 200 eligible and evaluable patients with 100 randomized into each treatment arm. The sample size is based on the following assumptions:

1. the median survival times for docetaxel and best supportive care are 7 and 4 months respectively
2. the patient accrual time is 10 months
3. the follow-up time after the last patient accrued is 5 months

Given the above assumptions, the sample size of 100 patients in each arm will result in an alpha level of 0.05 (2-sided) and power of 90% based on the log-rank test.

Interim Analysis

An interim analysis was planned after 50% of the patients had completed 6 cycles of docetaxel or discontinued the study. For this analysis the primary endpoint was survival time. Using the O'Brien-Fleming method, the log-rank test must be significant at 0.005 for the interim analysis and at 0.047 for the final analysis to be considered statistically significant.

Efficacy Endpoints

The primary efficacy endpoint is survival time, which is measured from the randomization date to the date of death from any cause or the date of last contact if there was no documentation of death.

The secondary efficacy endpoints include percentage of patients with objective response rate, duration of tumor response, time to progression, and quality of life measurements. Time to progression is defined from the randomization date to the date of first determination of progressive disease or to the date of the last assessment prior to further antitumor therapy (including radiotherapy). Quality of life measurements are obtained from the Lung Cancer Symptom Scale (LCSS) for North American centers and the EORTC Quality of Life Questionnaire for Lung Cancer for European centers. The endpoints defined from the LCSS are changes in LCSS scores from baseline scores rated on a 100 mm visual analog scale, and changes in ECOG performance status, body weight, and analgesic use.

Statistical Methods (Protocol)

1. The logrank and Kaplan Meier methods were used to analyze survival and time to progression. The median and Kaplan-Meier estimates were given with their 95% confidence intervals. Comparison between treatment groups was performed using the logrank test stratifying on ECOG performance status and response to prior platinum therapy.

2. Chi-square tests were used to compare response rates (replaced by Fisher's exact test if the expected frequency in one cell of the table is < 5). The 95% confidence interval for proportions was calculated following the exact method.

3. For the QOL endpoints, the sponsor proposed a missing data imputation strategy. Then, the QOL endpoints were analyzed using a longitudinal model and the GEE approach to test whether or not the assumption of randomly missing information holds. If the assumption does hold, then an analysis of covariance will be employed for the endpoint analysis using the last value carried forward method (LVCF) on the total scores and the subscale scores, while using a set of prospectively defined covariates. Each instrument will be analyzed separately. Then, the two instruments will be mapped against each other, utilizing the data

collected from a small portion of sites that will assess QOL on both the LCSS and EORTC QLQ-C30. Finally, if there are any merits, the total score as well as each-subscale score of each instrument will be transformed into a normal scale generating z-scores. These scores will also be compared collectively between the treatment groups.

Study Patients

The study population consists of 204 patients from both U.S. and non-U.S. sites. These patients constitute the intent to treat (ITT) population, which are all patients randomized. The first 100 patients were randomized to receive either 100mg/m^2 of docetaxel or best supportive care. These results were submitted in an interim analysis. The second 104 patients were randomized to receive either 75mg/m^2 or best supportive care. These results were compared separately from the first 100 patients.

The ITT population for time to progression includes randomized patients diagnosed with NSCLC since progression due to NSCLC could not be determined in patients not diagnosed with NSCLC. For time to progression, there are 202 patients. Two patients, 1 in each group, were excluded because they did not have NSCLC. The sponsor also defined an evaluable patient subgroup consisting of 192 patients, who were evaluable for efficacy.

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

The sponsor analyzed the primary and secondary endpoints separately for two time periods: those patients in the interim analysis (n=100) and those enrolled subsequently (n=104). The sponsor refers to these time periods as 317A and 317B. The patients in the docetaxel arm of 317A received 100 mg/m^2 of docetaxel, and patients in the docetaxel arm of 317B received 75 mg/m^2 of docetaxel. Among the patients in 317A, there were 49 patients in the docetaxel group and 51 in the BSC group. Among the patients in 317B, there were 55 patients in the docetaxel group and 49 in the BSC group. The patient population was fairly well-balanced for baseline and demographic characteristics within each time period but was not balanced with respect to baseline and demographic characteristics between these two periods. The sponsor showed using Fisher's Exact test for categorical variables and the two sample t-test for continuous variables that there were statistically significant differences between baseline and demographic characteristics between the two periods for age ($p < 0.01$), ECOG performance score ($p=0.04$), histological subtypes ($p=0.017$), and number of organs involved ($p=0.02$). In addition, of the patients who received radiation to the chest, there was statistically significantly less prior radiation to the chest ($\geq$ 50 GY vs. < 50 GY) in 317B patients than in 317A patients ($p<0.01$).

Survival Analysis

For study 317, overall survival was defined as the duration from the date of randomization to the date of death from any cause. Patients were censored at the date of last contact if there was no documentation of death, at the cut-off date of 4/12/99 if death occurred after the cutoff, or at the date of last assessment before the start of further anticancer therapy. Overall survival was analyzed using the Kaplan Meier method and the logrank test. The test did not reveal a statistically significant difference ($p=0.14$) between the treatment groups. The sponsor also analyzed the data separately in terms of the two time periods 317A and 317B. The results from Reviewer's Table 9 below show that there is a statistically significant survival difference in favor of the docetaxel arm in the 317B subgroup ($p=0.016$) only. The sponsor also reported that there was a statistically significant difference between the Kaplan Meier 1-year survival estimates in 317B, 40% in the docetaxel arm and 16% in the BSC arm (chi-square - $p=0.016$).

Reviewer's Table 9. Sponsor's Survival Analysis (ITT)

Cut-off Date: 4/12/99

	317A		317B		317	
	D-100	B-100	D-75	B-75	D-75	B-75
Patients randomized	49	51	55	49	104	100
Number of deaths(%)	42 (86.0)	40 (78.0)	35 (64.0)	35 (71.0)	77 (74.0)	75 (75.0)
Number censored(%)	7 (14.0)	11 (22.0)	20 (36.0)	14 (29.0)	27 (26.0)	25 (25.0)
Median survival (months)	5.9	4.9	9.0	4.6	7.2	4.7
95% CI (months)	4.5, 8.0	3.5, 8.0	5.5, 13.1	3.7, 6.1	5.5,9.2	3.7,6.0
1 yr. Survival K-M estimate (%)	19	28	40	16	28	23
95% CI (%)	7, 30	14, 41	26, 54	3, 30	19,38	13,32
Logrank test		p=0.871		p=0.016		p=0.14

The sponsor submitted an updated survival analysis in a **Fax dated November 5, 1999.** The cut-off date for the analysis was October 1, 1999. The following reviewer's table summarizes the sponsor's results based on the updated survival analysis.

Reviewer's Table 10. Sponsor's Updated Survival Analysis (ITT)

Cut-off Date: 10/1/99

	317A		317B		317	
	D-100	B-100	D-75	B-75	D-75	B-75
Median survival (months)	5.9	4.9	7.5	4.6	7.0	4.6
95% CI (months)	4.5, 8.0	3.5, 8.0	5.5, 12.8	3.7, 6.1	5.5,9.0	3.7,6.0
Logrank test		p=0.78		p=0.01		p=0.047

Reviewer's Comments:

1. The primary analysis should be based on data collected from **the pre-specified overall study period**. It is not appropriate to claim efficacy from a subgroup which was selected retrospectively. The 100 patients analyzed in the first study period actually comprise the patient population in the interim analysis. Therefore, the results of the final analysis should be based on the entire ITT population, not just the 104 patients subsequently enrolled. If the sponsor's purpose was to claim the efficacy for docetaxel 75 mg/m^2 group only after the interim analysis, the study design (e.g., type I error allocation, sample size recalculation with a given power level, statistical analysis plan, etc.) should be modified appropriately in the protocol.
2. The sponsor's updated survival analysis provided supportive evidence that patients treated with docetaxel may have better survival than those treated with best supportive care.

However, because the updated survival analysis is not a planned analysis, the interpretation of the p-value should be with caution.

3. The sponsor's chi-square method used to test for a difference between 1-year survival rates is valid, however, because the time point of 1 year was selected retrospectively, the statistically significant difference in the 1-year survival rates should be interpreted with caution. A more conservative approach, the logrank test, can be used while censoring patients who survive beyond the 1 year time point. The results of this analysis are given in Reviewer's Table 11. The logrank test gives a p-value of 0.08, which suggests that the difference in 1-year survival is not significant. The results of the 1-year survival analysis should be considered exploratory because the analysis was defined retrospectively.

Reviewer's Table 11. Reviewer's Survival Analysis to Test for Treatment Difference at 1 Year Survival (ITT), Cut-off Date: 4/12/99

	Docetaxel	BSC
Patients randomized	104	100
Number of deaths	68 (65.4%)	69 (69%)
Number censored	36 (34.6%)	31 (31%)
Logrank test vs. comparator	0.08	

4. In addition, if patients who died were counted as events at their date of death, and patients who didn't die were censored at their last follow-up date then, the result of logrank test does not give as strong evidence of a statistically significant survival difference in the 317B subgroup (p=0.041).

VII. Sponsor's Secondary Efficacy Results

Response Rate

Overall response rate is the sum of all confirmed complete and partial (CR + PR) responses in each treatment group. In the ITT population, the sponsor reported response rates of 5.8% and 0% in the docetaxel arm and the BSC arm respectively. There is a statistically significant difference (p=0.03) between the two arms.

Time to Progression

Time to progression (TTP) was analyzed for each treatment group using the Kaplan-Meier method. TTP was defined as the duration from the date of randomization to the date of documented progressive disease or death. Patients were censored at the date of last assessment if there was no documentation of progression, or at the date of last assessment before the start of further anticancer therapy. The results are summarized in Reviewer's Table 12.

Reviewer's Table 12. Sponsor's Time to Progression Analysis (ITT) – Censored at Subsequent Chemotherapy

	Docetaxel	BSC
Patients randomized	103	99
Number of PD	91 (88.3%)	82 (82.8%)
Number censored	12 (11.7%)	17 (17.2%)
Median time to PD (weeks)	10.6	6.7
95% CI (weeks)	7.6, 12.1	6.0, 7.3
26 week non-PD estimate (%)	16	5
95% CI (%)	8, 23	0, 10
Logrank test vs. comparator	p<0.001	

The analyses from Reviewer's Table 11 shows that there is a statistically significant difference between the docetaxel and the BSC groups ($p < 0.001$).

The MO did not believe that it was appropriate to censor patients at their time of further anticancer therapy, but instead, should have been counted as events at this time. In an exploratory analysis, these changes were made and the results presented in Table 13.

Reviewer's Table 13. Reviewer's Time to Progression Analysis (ITT)

	Docetaxel	BSC
Patients randomized	103	99
Number of PD	93 (90.3%)	85 (85.9%)
Number censored	10 (9.7%)	14 (14.1%)
Median time to PD (weeks)	10.1	6.8
95% CI (weeks)	7.6, 12.1	6.0, 8.0
26 week non-PD estimate (%)	16	7
95% CI (%)	9, 23	1, 13
Logrank test vs. comparator	0.008	

The results from Reviewer's Table 13 do not change much from the sponsor's results.

VIII. Quality of Life

The sponsor analyzed the QOL data for both studies by utilizing an analysis of covariance model (ANCOVA) on those patients who had paired QOL assessments at baseline and the last available assessment. The Lung Cancer Symptom Scale (LCSS) was comprised of two parts: the Patient scale and the Observer scale. The Patient scale included the following subscales: shortness of breath, cough, blood in the sputum, fatigue, appetite, pain, normal activities, lung cancer symptoms, and quality of life today. The Observer scale included the following subscales: loss of appetite, fatigue, cough, dyspnea, hemoptosis, and pain. All individual subscales were analyzed. Aggregrate scores were also analyzed. These included the Patient total scores, which consisted of adding the scores from the nine patient subscales, and the Observer total scores, which consisted of adding the scores from the six subscales. Additional aggregate scores in the Patient scale included Factor 1, Factor 1A, and Factor 2. Factor 1 was comprised of adding scores from the subscales fatigue, appetite, pain, normal activities, symptoms, and QOL today.

Factor 1A consisted of scores from the subscales fatigue, appetite and pain. Factor 2 was comprised of the scores from the subscales shortness of breath, coughing, and blood in the sputum. These three factors were defined as a result of performing a factor analysis on all of the subscales. Up to 3 questions were allowed to be missing in order to come up with a score, otherwise the score was considered missing. Missing values were imputed. Comparisons were undertaken at baseline and the last available assessment. These results are shown in the sponsor's Tables 4.07D in the appendix.

Reviewer's Comment: P-values provided in the sponsor's Tables 4.07D should be interpreted with caution because of multiplicity. This approach ignores the impact of missing data, which may very likely be informative. Serious bias can occur if missing data are related to a nonrandom mechanism. In addition, the use of imputation in computing aggregate scores presumes that each question has equal weight in determining clinical benefit, which may not be the case. The clinical interpretation of these aggregate scores is also in question. The sponsor was unable to externally validate the results of the factor analysis used to justify the definition of Factors 1, 1A, 2. Therefore, it is not appropriate to include these as measures of quality of life.

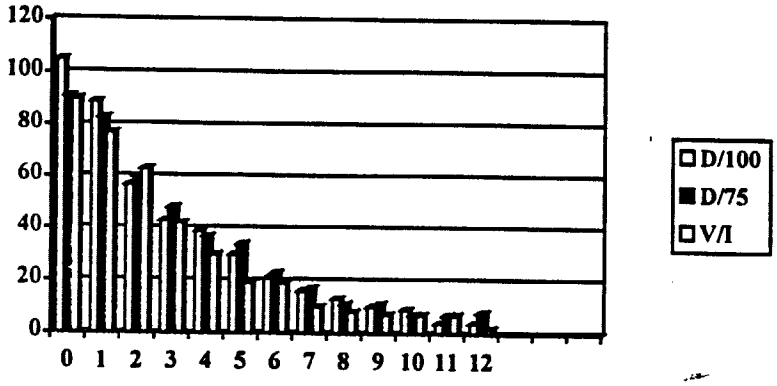
Tax 320

The sponsor also performed a longitudinal data analysis using a pattern mixture model. The analysis was performed using change from baseline on all available assessments during the study including the follow-up period up to and including 259 days after the first infusion (first 12 periods). This analysis was performed on each individual and aggregate score. The missingness issue was analyzed by dividing patients into non-completers and completers. The non-completers were those patients who were treated only up to cycle 3, and completers were those patients who were treated for more than 3 cycles. There were 144 (56 – D/100, 40 – D/75, 48 – V/I) patients defined as non-completers and 142 (49 – D/100, 51 – D/75, 42 V/I) patients defined as completers. The results of this analysis for completers are shown in the sponsor's Table 105 in the appendix. From this table, none of the subscales are shown to have a statistically significant treatment effect between either docetaxel arm or the control arm.

The MO believed certain questions in the questionnaire had greater clinical relevance than others. The questions selected were Patient reported pain and QOL today scores. For all QOL parameters, the higher the score reported, the "better" the patient felt. QOL data was analyzed for the ITT population.

To analyze the 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 each of the first 12 periods starting from baseline. From Figure 1 it is evident that throughout the treatment and follow-up periods, all 3 arms display similar rates of attrition. Attrition is especially high after the third period in all 3 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 three treatment groups. This reviewer employed a growth curve analysis, using period as the unit of time, where period is defined as a three-week time interval. An estimated time trend describes the effect of treatment over the study period, where time trend was analyzed by means of a GEE linear model. Since the response is defined as the change from baseline on all available assessments, the models were fit with only a slope term. 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.

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 third period and those who continued to be assessed beyond the third period. The former group is classified as "dropouts" and the latter "completers". This cutoff was selected by the sponsor because all treatment arms experienced a 50% dropout rate by this period. There are 154 patients (62 – D/100, 43 – D/75, 49 – V/I) denoted as dropouts and 132 patients (43 – D/100, 48 – D/75, 41 – V/I) denoted as completers. In Reviewer's Tables 14 and 15 the results are presented for dropouts and completers.

Reviewer's Table 14. QOL Analyses for Dropouts

Treatment	Parameter	Estimate	St Error *	P-value
D/100	Pain	1.934	2.630	0.4686
D/100	QOL today	1.421	2.567	0.5846
D/75	Pain	-6.145	2.591	0.0279
D/75	QOL today	1.299	2.632	0.6272
V/I	Pain	-1.378	2.243	0.5429
V/I	QOL today	-4.197	1.973	0.0408

* Robust standard errors are provided.

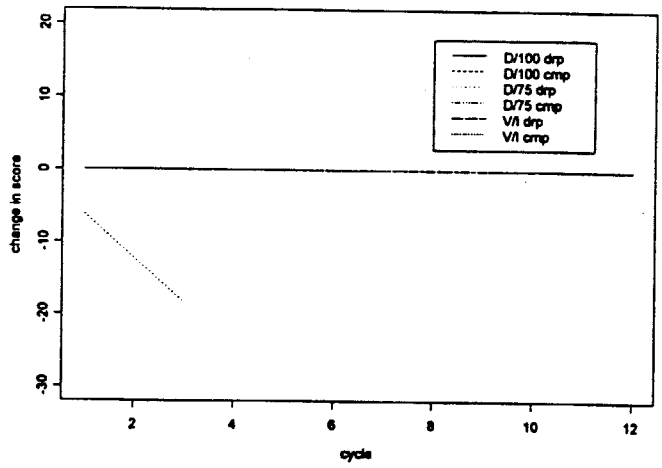
Reviewer's Table 15. QOL Analyses for Completers

Treatment	Parameter	Estimate	St Error *	P-value
D/100	Pain	-0.605	0.383	0.1156
D/100	QOL today	-1.185	0.342	0.0007
D/75	Pain	-0.271	0.338	0.4243
D/75	QOL today	-0.868	0.354	0.0151
V/I	Pain	-0.830	0.428	0.0544
V/I	QOL today	-1.556	0.366	< 0.0001

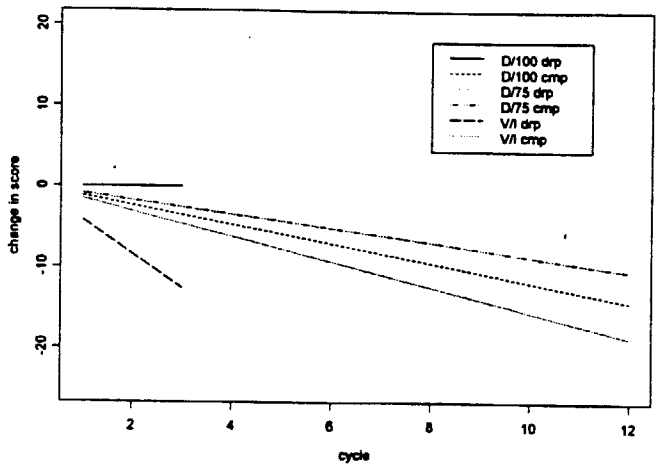
Among dropouts, the estimates of the D/100 arm and the V/I arm for the pain parameter are not statistically significant. Therefore, there is a constant time trend for these two treatment arms. The estimate for the D/75 arm, however, is statistically signifcant and shows a negative linear trend. For the QOL today parameter, the estimates for the docetaxel arms are not statistically significant. As a result, these two treatment arms show a constant time trend. On the other hand, the estimate for the V/I arm is statistically significant and shows a negative linear trend.

Among completers, the slope estimates of all three treatment arms for pain are not statistically significant, thus suggesting a constant trend over time. However, for the parameter QOL today, all three treatment arms have statistically significant slope estimates and, therefore, display different negative linear trends. These results are depicted in Reviewer's Figures 2 and 3. Figure 2 shows that for the pain parameter, the time trends are constant for D/100 dropouts and completers and for V/I dropouts and completers.

Reviewer's Figure 2. Change in Pain Scores for Completer and Dropout Groups



Reviewer's Figure 3. Change in QOL Today for Completer and Dropout Groups



A test for non-ignorable missing can be performed within each treatment arm. This test consists of computing a t-statistic based on the difference between the estimates for the dropout and completer groups and a pooled standard error. Results are shown in Reviewer's Table 16.

Reviewer's Table 16. Results for Test of Informative Missing

QOL Parameter	Treatment group	p-value
Pain	D/100	0.0740
Pain	D/75	<0.0001
Pain	V/I	0.6928
QOL today	D/100	0.0492
QOL today	D/75	0.1332
QOL today	V/I	0.0293

For the pain parameter, the test for non-ignorable missing reveals that for the D/100 and the V/I groups, the missing mechanism can be assumed to be noninformative. As a result, analysis of the pain parameter should not be biased as a result of dropouts. However, this is not the case with the D/75 arm. The time trends are clearly different for dropouts and completers. Dropouts experience an increase in pain as shown by the sharp decline, and completers experience no change in their pain over time. In addition, in the D/75 arm, the test for non-ignorable missing shows that missing data is informative, and therefore, dropouts and completers need to be analyzed separately For the QOL today parameter, Figure 4 shows that the time trends are negatively sloped fc all six treatment and status groups, but there do not appear to be very large differences between them. The test for non-ignorable missing reveals that for the D/100 and V/I treatment groups, missing information cannot be considered ignorable so that dropouts and completers need to be analyzed separately. However, in the D/75 treatment group, missing information can be considered noninformative. As a result, it is possible to combine dropouts and completers in the D/75 group but not the D/100 or V/I groups.

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 and a pooled standard error. Reviewer's Table 17 shows the p-values from performing such tests:

Reviewer's Table 17. Results for Test of Treatment Difference

QOL Parameter	D/100 vs. V/I	D/75 vs. V/I
Pain (drop)	0.3410	0.1860
Pain (comp)	0.6952	0.3011
QOL today (drop)	0.0829	0.1012
QOL today (comp)	0.4614	0.1893

Results in the above table demonstrate that the estimated time trends are similar between the D/100 vs. the V/I arms and between the D/75 vs. the V/I arms within dropouts and completers. According to these results, it appears that there is not much long term clinical benefit in the either docetaxel group.

This reviewer could not find a statistically significant treatment difference between the docetaxel and control groups. However, patients in the docetaxel arms do not experience a change in their pain scores as shown in Reviewer's Figure 2. With respect to QOL today (Reviewer's Figure 3), patients in the docetaxel arms show slower rates of decline with the slowest rate being the D/75 group. Based on these observations, it appears that the quality of life for the docetaxel group is not worse than that of the control group. A possible reason for the inability to show a statistically significant difference may be due to the high attrition rates in all three arms.

Sponsor's Exploratory QOL Analysis in TAX 320

The sponsor performed exploratory QOL analyses in an attempt to correlate responses to the QOL data. One of the analyses performed was to measure the mean changes in Patient Total Score from baseline to one of the following time points: the best score on treatment or follow-up, period 1, period 2, or period 3 in subgroups defined by the best response on study. These mean changes are summarized in the Sponsor's Table 106. The sponsor stated that in certain subgroups of patients, one or both of the docetaxel groups showed favorable changes from baseline compared to the control group.

Reviewer's Comment: It is not appropriate to divide the population into responders and non-responders because there may be other factors that may account for these mean differences other than treatment.

Tax 317

Quality of life was measured using both the LCSS and the EORTC QLQ-C30 and QLQ-LC13 instruments. Because very few patients filled out the EORTC QLQ-C30 and QLQ-LC13 questionnaires, only the data from the LCSS instrument will be reviewed.

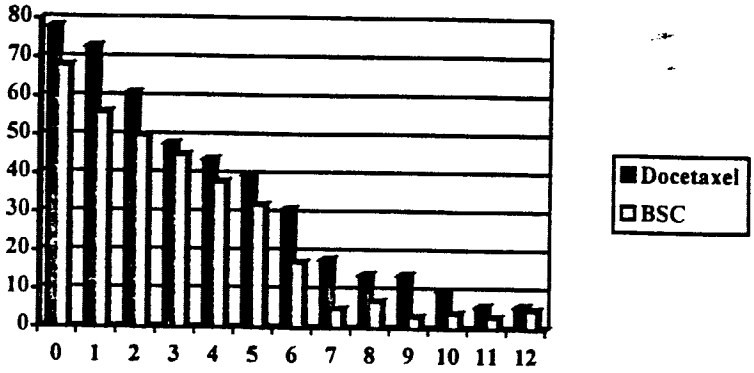
Results are shown in the sponsor's Table 4.07D in the appendix. (See description of sponsor's analyses and Reviewer's comments from TAX 320).

The sponsor also performed a longitudinal data analysis using a pattern mixture model. The analysis was performed using change from baseline on all available assessments during the study including the follow-up period up to and including 259 days after the first infusion (first 12 periods). This analysis was performed on each individual and aggregate score. The missingness issue was analyzed by dividing patients into non-completers and completers. The non-completers were those patients who had their last QOL assessment prior to or in period 3 and completers were those patients who had their last QOL assessment after period 3. The results of

this analysis for completers are shown in the sponsor's Table 51 in the appendix. From this table, only the pain subscale is shown to have a statistically significant treatment effect between the docetaxel arm and the control arm.

Reviewer's Figure 4 shows the frequencies of patients at each of the first 12 periods starting from baseline. From Figure 4 it is evident that throughout the treatment and follow-up periods, both arms display similar rates of attrition. Attrition is especially high after the third period in both arms.

Reviewer's Figure 4. Number of Patients over Time



Using a pattern mixture model (Little, 1993 and 1995), patients are divided into two groups: those who dropped out before the third period and those who continued to be assessed beyond the third period. This cutoff was selected by the sponsor because all treatment arms experienced a 50% dropout rate by this period. There are 48 patients (26 – Docetaxel, 22 – BSC) denoted as dropouts and 98 patients (52 – Docetaxel, 46 – BSC) denoted as completers. In Reviewer's Tables 18 and 19 the results are presented for dropouts and completers.

Reviewer's Table 18. QOL Analyses for Dropouts

Treatment	Parameter	Estimate	St Error *	P-value
Docetaxel	Pain	5.150	2.376	0.0438
Docetaxel	QOL today	-4.413	2.935	0.1510
BSC	Pain	-0.578	2.528	0.8217
BSC	QOL today	-2.346	2.853	0.4230

* Robust standard errors are provided.

Reviewer's Table 19. QOL Analyses for Completers

Treatment	Parameter	Estimate	St Error *	P-value
Docetaxel	Pain	-0.864	0.298	0.0040
Docetaxel	QOL today	-1.399	0.358	0.0001
BSC	Pain	-1.634	0.558	0.0038
BSC	QOL today	-2.022	0.619	0.0013

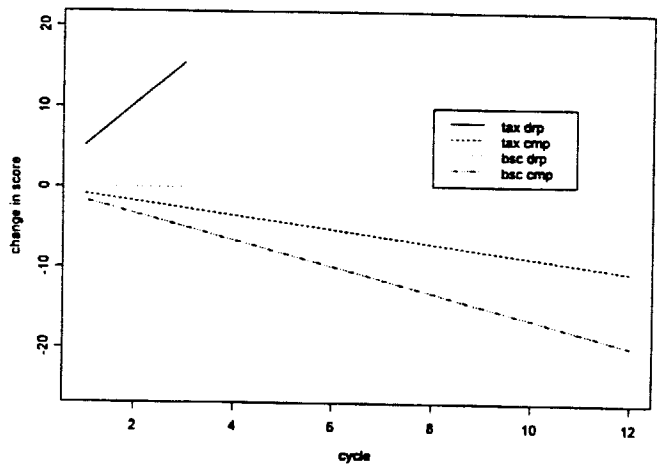
* Robust standard errors are provided.

Among dropouts, the estimate of the docetaxel arm for the pain parameter is statistically significant; however, the estimate of the BSC arm is not. Therefore, there is a positive linear trend for the docetaxel arm, and a constant time trend for the BSC arm. For the QOL today parameter, neither the estimates of the docetaxel arm nor the BSC arm are statistically significant. As a result, these two treatment arms show a constant time trend.

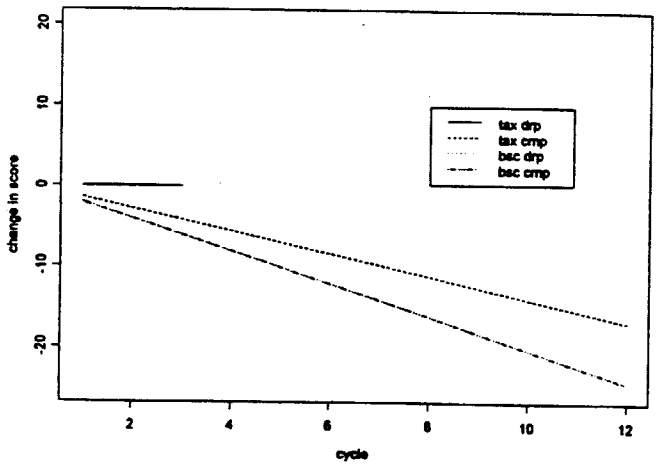
Among completers, the slope estimates of both treatment arms for pain are statistically significant, which suggests both arms show negative linear trends. For the parameter QOL today, both treatment arms have statistically significant slope estimates and display negative linear trends.

These results are depicted in Reviewer's Figures 5 and 6. Figure 5 shows that for the pain parameter, the time trend is positive for docetaxel dropouts and negative for docetaxel completers. For the BSC arm, dropouts show no change in pain response, while completers show a negative trend. A test for non-ignorable missing gives the following results as shown in Reviewer's Table 20.

Reviewer's Figure 5. Change in Pain Scores for Completer and Dropout Groups



Reviewer's Figure 6. Change in QOL Today Scores for Completer and Dropout Groups



Reviewer's Table 20. Results for Test of Informative Missing

QOL Parameter	Treatment group	p-value
Pain	Docetaxel	<0.0001
Pain	BSC	0.5907
QOL today	Docetaxel	0.0587
QOL today	BSC	0.8838

For the pain parameter, the test for non-ignorable missing revealed that for the docetaxel group, the missing mechanism is assumed to be informative. Therefore, dropouts and completers need to be analyzed separately. Clearly, the time trends are different for these two groups. Dropouts in the docetaxel group experience a decrease in pain over time as shown by the positive slope. Completers, on the other hand, experience an increase in pain over time as shown by the negative slope. In the BSC arm, the test for non-ignorable missing reveals that the missing mechanism can be assumed to be noninformative. As a result, analysis of the pain parameter should not be biased as a result of dropouts. For the QOL today parameter, Figure 6 shows that the time trends are negatively sloped for completers from both treatment groups, but there is no change in QOL today for both dropout groups. The test for non-ignorable missing reveals that for both treatment groups, missing information can be considered ignorable so that dropouts and completers can be combined.

A test for difference in treatment arms can be performed with results shown in Reviewer's Table 21.

Reviewer's Table 21. Results for Test of Treatment Difference

QOL Parameter	Docetaxel vs. BSC
Pain (drop)	0.1079
Pain (comp)	0.1874
QOL today (drop)	0.6178
QOL today (comp)	0.3517

Results in the above table demonstrate that the estimated time trends are similar between the two arms within dropouts and completers. According to these results, it appears that there is not much long term clinical benefit in the docetaxel group.

To summarize the results in TAX 317, a statistically significant treatment difference between the docetaxel and the BSC treatment arms for the pain parameter could not be confirmed by this reviewer. However, from Reviewer's Figure 5, it appears that patients in the docetaxel arm experience less pain over time (dropouts) or do not decline as quickly as patients in the BSC arm (completers). For QOL today (Reviewer's Figure 6), patients in the docetaxel group either experience no change in their QOL scores (dropouts) or experience slower rates of decline compared to the BSC group. Based on these observations, it appears that the quality of life for the docetaxel group is not worse than that of the control group. A possible reason for the inability to show a statistically significant difference may be due to the high attrition rates in both arms.

In addition, the LCSS questions may be subjective and answers can vary depending on an individual patient's perception. For example, for the question of how much pain one has, the responder is asked to mark on a scale of ranging from "none" to "as much as it could be". The

answer may depend on the patient's perception of pain and ability to tolerate it. Thus, the subjectivity issue in this unblinded trial setting is another factor complicating the interpretation of the QOL findings.

IX. Exploratory Survival Analyses on TAX 320 and TAX 317

This reviewer analyzed the survival data on the D/75 arms and the corresponding control groups for the pivotal studies to determine if the prospectively defined prognostic factors on survival time influenced the treatment effect. Thus, for TAX 320, the analysis population was comprised of the 125 patients in the D/75 arm and the 123 patients in the V/I arm. For TAX 317, the analysis population consisted of the 55 patients in the D/75 arm and the 49 patients in the B/75 arm. The prognostic factors specified were performance status, response to prior platinum therapy, age, disease stage at study entry, weight loss, prior radiation therapy, number of prior chemotherapy regimens, overall exposure to platinum, level of LDH, sex, and country (TAX 317 only). LDH level was excluded because many patients did not have baseline LDH level. Cox models, adjusting for all prospectively defined factors, were fitted. Factors that were not significant were eliminated. The final models are given in Reviewer's Tables 22 and 23.

Reviewer's Table 22. Final Cox Model for TAX 320

Prognostic Factor	Risk Ratio	p-value
Performance Status (0,1)	0.303	0.0001
Disease stage at study entry (IV)	1.894	0.0222
Weight loss (< 10%)	0.462	0.0013
Prior radiotherapy	1.396	0.0315
Treatment (D/75)	**0.853**	**0.2677**

Reviewer's Table 23. Final Cox Model for TAX 317

Prognostic Factor	Risk Ratio	p-value
Performance Status (0,1)	0.356	0.0003
Disease stage at study entry (IV)	3.116	0.0011
Treatment (D/75)	**0.530**	**0.0108**

Based on results from these tables, the treatment effect is not significant in the TAX 320 study. However, it is significant in the TAX 317 study. These results confirm those of the unadjusted analyses.

X. Overall Summary

TAX 320 and TAX 317 were multi-center, open-label, randomized Phase III trials. The objective of TAX 320 was to compare survival in patients with non-small cell lung cancer (NSCLC) previously treated with platinum-containing chemotherapy receiving 100 or 75 mg/m^2 docetaxel vs. either vinorelbine 30mg/m^2 per week or ifosfamide 6 mg/m^2. The objective of TAX 317 was to compare survival in patients with non-small cell lung cancer (NSCLC) previously treated with platinum-containing chemotherapy receiving either docetaxel (100 mg/m^2, then 75 mg/m^2) or best supportive care. The statistical review can be summarized as follows:

1. In study TAX 320, a statistically significant survival difference was not demonstrated between the two docetaxel arms and the control group. In study TAX 317, based on a protocol specified cut-off date of April 12, 1999, a statistically significant survival

difference was not demonstrated in the ITT population, but was shown in the subgroup (defined retrospectively) comprised of the 104 patients after the dose reduction to 75 mg/m^2. In the updated survival analysis with the cutoff date at October 1, 1999, a significant survival difference favoring the docetaxel arm (p=0.047) was shown in the pooled analysis (100 and 75 mg/m^2). However, because this updated survival analysis was an unplanned analysis, the p-value should be interpreted with caution. It should be noted the patient population in TAX 317 was comprised of patients from both U.S. and non-U.S. sites.

2. The response rate was statistically significantly greater in the docetaxel arms when compared to the control arm in TAX 320. This result was replicated in TAX 317.

3. In TAX 320, a statistically significant difference in TTP was not shown. However, in TAX 317, a statistically significant difference in TTP was demonstrated in favor of the docetaxel group.

4. In the analysis of the QOL parameters for both studies, a statistically significant difference between the docetaxel and control groups could not be demonstrated.

XI. Conclusions and Recommendations

In this supplemental NDA submission the primary efficacy endpoint is survival time. The study does not provide sufficient evidence to support the conclusion that docetaxel offers a treatment benefit in patients with NSCLC previously treated with platinum-based chemotherapy. The sponsor's claim of efficacy based on the subgroup of TAX 317 taking the 75 mg/m^2 dose is in question since this result is from an unplanned analysis.

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, it appears that patients in the docetaxel groups did not do worse with respect to their quality of life than patients in the control groups.

/S/

Clara Chu, Ph.D
Mathematical Statistician

Concur: Dr. Chen /S/ 12/15/99
Dr. Chi

C:\My Documents\review\taxotere\tax_rev.doc
This review consists of 23 pages of text.

References

Beitz J., Gnecco C., Justice R. "Quality-of-Life End Points in Cancer Clinical Trials: The U.S. Food and Drug Administration Perspective," *Journal of the National Cancer Institute Monographs*, 20, 7-9, 1996.

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.

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