CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER: NDA 20-262/S-024

STATISTICAL REVIEW(S)

Statistical Review and Evaluation

MAR 30 1998

NDA Number:	20-262 /S-024
Applicant:	Bristol-Myers Squibb
Name of Drug:	Taxol (paclitaxel)
Indication:	Treatment of non-small cell lung cancer
Document Reviewed:	Vols. 8/10.1-8/10.8 submission dated June 30, 1997; Vol. 99.1 submission dated January 21, 1998
Medical Reviewer:	Isagani Chico, M.D.
Statistical Reviewer:	David Smith, Ph.D.

Contents

1 Background and Overview

In order to support labeling for the indication of treatment of advanced non-small lung cancer, the sponsor submitted a supplemental NDA which is comprised of 3 Phase III trials and 4 Phase II trials.

A brief summary of the studies appears below.

Study	Type	*N*	Arms
CA139-165	Randomized Ph III	599	Cisplatin/Etoposide
			Taxol/Cisplatin/G-CSF
			Taxol/Cisplatin
CA139-103	Randomized Ph II/III	332	Teniposide/Cisplatin
			Taxol/Cisplatin
CA139-208	Randomized Ph III	414	Cisplatin
			Taxol/Cisplatin
CA139-027	Non-randomized Ph II	27	Taxol
CA139-029	Randomized Ph II	118	Taxol
			Merbarone
			Piroxantrone
CA139-127	Non-randomized Ph II	21	Taxol
CA139-201	Non-randomized Ph II	58	Taxol

The sponsor's submission includes the reports of three pivotal studies (CA139-165, CA139-103, and CA139-208), as well as four Phase II supportive studies. We will only consider the three pivotal studies in this review. The next section includes important statistical issues in each study. The following sections consider each study separately and these will be self-contained. Each section will be divided into three subsections:

1. General description of the study
2. Efficacy endpoints and results
3. Summary and conclusions

The final section of this review will include overall conclusions and recommendations for the submission.

References and an appendix of statistical tables and figures follow the review.

2 Summary of Statistical Issues

2.1 Issues in CA139-165

- The sponsor performed a "pooled Taxol arm" analysis although it was not pre-specified in the protocol. That is, there were two distinct Taxol-containing regimens in this study, each with different doses. The sponsor pooled patients in both Taxol arms in order to compare survival on this "pooled" arm with the cisplatin/etoposide arm. The median survival was 7.4 months (95% 6.5-8.6 months) for the cisplatin/etoposide group and 9.7 months (95% 8.8-10.6 months) for the pooled taxol group (stratified log-rank p = 0.049). The sponsor concluded that there was a statistically significant difference. However, the nominal Type I error rate must be less than 0.0125 given the fact that this comparison is made in addition to the three comparisons that were pre-specified in the protocol (see Section 3.1). Therefore, one must conclude that there is no statistically significant difference between survival in the pooled Taxol group and the cisplatin/etoposide group.

- Seven patients were categorized as "having evidence of disease progression," although there was no apparent reason for their progression on their case-report forms. The medical reviewer and this reviewer analyzed the time to disease progression, changing their "progressed" status to "censored."

- The quality of life analysis consisted of comparing subscale measurements at each followup to baseline measurements using the Wei-Lachin statistic. The assumptions of this analysis rely heavily on the behavior of patient dropout. Namely, it assumes that patients are missing completely at random. The sponsor did not perform a longitudinal quality of life analysis, which is more appropriate in determining trends over time and variance inflation due to non-random patient dropout.

- The sponsor reported a statistically significant improvement of Taxol/cisplatin over cisplatin/etoposide in the Lung Cancer Symptoms subscale. The p-value corresponding to this hypothesis test was 0.027, but the nominal Type I error level for testing for an improvement is 0.0125. This departure from 0.05 is due to multiple treatment adjustments. The correct conclusion in

this case is that there is no statistically significant difference in Lung Cancer Symptoms between Taxol/cisplatin and cisplatin/etoposide.

- There was no Type I error adjustment for the number of secondary endpoints considered (including multiple QOL subscales), and therefore the p-values of each of these endpoints should be considered to be inflated.

2.2 Issues in CA139-103

- As in CA139-165, the QOL analysis consisted of comparing on-study subscale measurements to baseline measurements without testing whether patient dropout would bias the conclusions. The Wei-Johnson test that was used to compare QOL subscales assumes that dropout is completely random. A longitudinal approach would have been much more robust to assess dropout patterns.

- There was no Type I error adjustment for the number of secondary endpoints considered (including multiple QOL subscales), and therefore the p-values of each of these endpoints should be considered to be inflated.

2.3 Issues in CA139-208

- In the QOL analysis of this study, there was no hypothesis test for determining whether patient dropout was ignorable or nonignorable. Although the sponsor made an effort to examine informally how patient dropout would effect the QOL conclusions, there was no attempt to quantify this effect objectively.

- There was no Type I error adjustment for the number of secondary endpoints considered (including multiple QOL subscales), and therefore the p-values of each of these endpoints should be considered to be inflated.

3 Study CA139-165

3.1 Description of CA139-165

Study objective: To determine the duration of survival of patients with advanced

non-small cell lung cancer (NSCLC) in patients receiving treatment with Taxol and to determine the qualitative and quantitative toxicity of Taxol.

Study Enrollment Period: September 1993 - March 1994

Study Design: Open label, multi-center, randomized clinical trial. The stratification factors were performance status, weight loss in the previous six months, and disease stage.

Sample size: Of 599 patients randomized, 198 were treated with Taxol/cisplatin, 201 were treated with Taxol/cisplatin/G-CSF, and 200 were treated with cisplatin/etoposide.

The sponsor based sample size calculations on survival. Considering that there were three treatments in this trial, the sponsor accounted for multiple comparisons among the three treatments by evaluating Taxol/cisplatin vs. cisplatin/etoposide and Taxol/cisplatin/G-CSF vs. cisplatin/etoposide each with a one-sided Type I error of 0.0125. The remaining 0.025 Type I error was used for a two-sided evaluation of Taxol/cisplatin/G-CSF vs. Taxol/cisplatin. This guaranteed a global Type I error of 0.05.

For a one-sided Type I error of 0.0125, a total of at least 585 patients (195 per arm) will give an attainable power of at least 0.825 for detecting an increase of 50% in median survival in favor of Taxol/cisplatin over cisplatin/etoposide. The median survival for cisplatin/etoposide was assumed to be six months.

Interim analysis: None.

Dose: The Taxol/cisplatin treatment group received 135 mg/m^2 or 250 mg/m^2 of Taxol with G-CSF as a 24 hour infusion, followed by 75 mg/m^2 of cisplatin over 1 hour. This was repeated every 21 days.

The cisplatin/etoposide group received 75 mg/m^2 of cisplatin over 1 hour on day 1 and 100 mg/m^2/day of etoposide over 45 minutes, days 1, 2, and 3.

Treatment duration: Patients with a complete or partial response or stable disease were to be treated until evidence of disease progression. Patients were to be removed from the study for progressive disease or excessive toxicity, or they were removed after 6 courses if at least one of the following conditions existed: weight loss $\geq$ 5% of baseline weight, decrease in ECOG performance status of one level and Grade III nausea.

Criteria for Evaluation: The primary endpoint for comparison was survival; this was tested with a stratified log-rank test. Standard WHO response criteria were used (complete / partial remissions / stable disease) and common toxicity criteria (CTC) were used for toxicity. The other secondary efficacy endpoints considered were time to progression, objective response, duration of response, and quality of life (QOL).

3.2 Efficacy Endpoints and Results of CA139-165

Recall that comparisons between cisplatin/etoposide and either Taxol arm must be made at a nominal 0.0125 level instead of the usual 0.05 level in order to preserve Type I error across multiple comparisons.

Survival Analyses

Table 1 shows the overall survival estimates and hazard ratios for the three arms. The median survival time in the Taxol/cisplatin arm was 9.3 months (95% CI 8.0-10.4 months), the median survival time in the Taxol/cisplatin/G-CSF arm was 10.0 months (95% CI 8.9-11.7 months), and the median survival time in the cisplatin/etoposide arm was 7.4 months (95% CI 6.5-8.6 months). There was no statistically significant difference between Taxol/cisplatin and cisplatin/etoposide (stratified log-rank $p = 0.125$) or between Taxol/cisplatin/G-CSF and cisplatin/etoposide (stratified log-rank=0.079); the stratification factors were performance status, weight loss, and disease stage. There was no statistically significant difference between survival for Taxol/cisplatin and Taxol/cisplatin/G-CSF (stratified log-rank $p = 0.75$).

One year survival rates were 36% (95% CI 30 to 43%) for the Taxol/cisplatin arm, 40% (95% CI 34 to 47%) for the Taxol/cisplatin/G-CSF arm, and 32% (95% CI 26% to 39%) for the cisplatin/etoposide arm.

Given the fact that there was no statistically significant difference in survival between Taxol arms, the sponsor pooled the two Taxol arms and compared this pooled treatment group with the cisplatin/etoposide arm. This analysis was not pre-specified in the protocol. The results appear in Table 2. The median survival was 7.4 months (95% 6.5-8.6 months) for the cisplatin/etoposide group and 9.7 months (95% 8.8-10.6 months) for the pooled taxol group (stratified log-rank $p = 0.049$). The sponsor concluded that there was a statistically significant difference. However, the nominal Type I error rate must be less than 0.0125 given the fact

that this comparison is made in addition to the three comparisons that were pre-specified in the protocol (see Section 3.1). Therefore, one must conclude that there is no statistically significant difference between survival in the pooled Taxol group and the cisplatin/etoposide group.

Cox Regression Analysis on Survival Data

As a secondary analysis, the sponsor performed a Cox regression on survival time, pre-specifying the following covariates in the protocol: gender, LDH at baseline, prior radiotherapy (yes/no), and histology. The strata were performance status, weight loss, and disease stage. Table 3 shows the results of the Cox model on the survival data.

Note that neither Taxol treatment explains sufficient variability in survival to be statistically significant. The only statistically significant variable among the ones measured in Table 3 is LDH.

Time to Progression Analysis

Tables 4 and 5 show the frequencies of progression and time to progression (TTP) across the three treatments. The difference in TTP between Taxol/cisplatin and cisplatin/etoposide was not statistically significant ($p = 0.0504$ log-rank test), but there was a statistically significant difference between Taxol/cisplatin/G-CSF and cisplatin/etoposide ($p = 0.004$ log-rank test). Recall that the per-comparison level is 0.0125, not 0.05.

The sponsor performed a stratified Cox regression for time to progression. The results of this analysis appear in Table 6. Taxol arm vs. cisplatin/etoposide as a prognostic factor was statistically significant for Taxol/cisplatin/G-CSF but not statistically significant for Taxol/cisplatin ($p = 0.005$ for Taxol/cisplatin/G-CSF; $p = 0.060$ for Taxol/cisplatin). In addition, LDH also explained statistically significant variability in time to progression.

There were seven patients who were classified as having "tumor progression," but there was no corresponding evidence of progression on these patients' case-report forms. These patients' identifying numbers appear in Table 7. This reviewer performed an analysis of time to progression with these seven patients classified as "censored." The results of this analysis appears in Table 8. The conclusions do not

change from those indicated in Table 5.

Clinical Response

Patients in this study were required to have histologically confirmed non-small cell bronchogenic carcinoma (either stage IIIB or stage IV). Patients were also required to have either measurable (bidimensionally) or evaluable disease (unidimentionally measurable or non-measurable). The sponsor's response assessment was compared to the best assessment by the ECOG study chairman's.

Table 9 shows the tumor response rates across treatment groups. In an intent-to-treat (ITT) analysis, the proportion of objective response (CR+PR) for the cisplatin/etoposide arm was 12% (95% CI 8-17%), compared with 23% (95% CI 18-30%) in the Taxol/cisplatin arm and 25% (95% CI 20-32%) in the Taxol/cisplatin/G-CSF arm. The response rate advantage of Taxol/cisplatin over cisplatin/etoposide was statistically significant ($p = 0.001$). The response rate advantage of Taxol/cisplatin/G-CSF over cisplatin/etoposide was also statistically significant ($p < 0.001$).

Duration of Response

Duration of response was calculated for all responders (CR+PR) and was defined as the period from the first day of treatment until the date progression was first noticed. Eighty-four percent (40/46) of Taxol/cisplatin responders progressed or experienced disease recurrence. The median duration of response for the Taxol/cisplatin arm was 7.1 months with range 3.3 to 29.4+ months. Eighty percent (41/51) of the Taxol/cisplatin/G-CSF responders progressed and the median duration of response was 9.2 months with a range of 2.2+ to 24.7 months. Eighty-eight percent of the responders in the cisplatin/etoposide group progressed. The median duration of response for the cisplatin/etoposide arm was 7.5 months with a range of 3.0 to 24.7 months.

There was no statistically significant difference between cisplatin/etoposide and Taxol/cisplatin (stratified log-rank $p = 0.681$) or Taxol/cisplatin/G-CSF (stratified log-rank $p = 0.227$) with respect to duration of response.

Quality of Life

The sponsor administered evaluated quality of life (QOL) based on the FACT-G and FACT-L questionnaire. Table 10 shows the questionnaire compliance across the three arms with respect to followup.

The sponsor compared QOL across treatment arms by examining median change at each time point to baseline. This was tested by using the Wei-Lachin test for stochastic ordering. The sponsor found no statistically significant differences between the Taxol arms versus the cisplatin/etoposide arms in the seven subscales examined, except for an improvement of Lung Cancer Symptoms from baseline when comparing Taxol/cisplatin to cisplatin/etoposide. This, however, is not the correct conclusion to draw from the Lung Cancer Symptoms p-value of 0.027. Since the nominal Type I error level is 0.0125 (and not 0.05) due to the nature of multiple comparisons among three treatment arms, it must be concluded that there is no statistically significant difference in the Lung Cancer Symptoms subscale.

This reviewer had several concerns with respect to this analysis. The sponsor reported no method of controlling Type I error to account for the number of QOL subscales that were considered. Also, the Wei-Lachin test assumes that dropout is completely at random, which is not always the case in cancer trials.

Fairclough (1997) performed a longitudinal analysis of QOL data on this data set. This reviewer will not report in depth the methods or conclusions of this analysis; however, one primary objective of this study was to compare total QOL subscale score among the treatment arms, which is relevant to this review. The main conclusion of Fairclough (1997) was that "health-related" QOL strongly predicts response to therapy and survival among these patients (as evidenced by QOL factors explaining large amounts of variability of survival in proportional hazards models). However, there were no statistically significant differences between treatment arms with respect to change in QOL endpoints over time.

Analyzing QOL data presents us with two challenges. The first challenge is that of within-patient correlation across the repeatedly measured QOL endpoints. The second challenge is that of missing data. Implicitly, the sponsor assumed that dropout in CA139-165 was completely random across arms. However, this is not always the case, particularly in cancer trials.

In a classical univariate repeated ANOVA, a particular correlation structure known as compound symmetry must be assumed for a valid F-test of interaction of treatment and time. A multivariate approach may be considered when a compound

symmetry assumption fails. However, in a multivariate approach, a distribution must be explicitly specified with the "correct" mean and covariance matrix.

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 the under-estimation of the covariance matrix by introducing "sandwich" estimators. Therefore, we have some assurance of a variance estimate that is robust.

This reviewer analyzed the QOL data using a GEE linear model and derived a robust covariance estimator based on M-estimation theory. To deal with the problem of potentially informative dropout, this reviewer based the dropout analyses on the concept of a pattern-mixture model (Little, 1993 and 1995).

Only three QOL subscales were considered in this reviewer's analysis. These were Physical Functioning, Functional Well-Being, and Lung Cancer Specific Symptoms. These were determined to be the most clinically relevant QOL measurements by both this reviewer and the medical reviewer.

Table 11 shows the results of fitting the models described above to this study's QOL data. It was determined from graphical methods that, in the cisplatin/etoposide arms, the estimated linear trends for all three subscales were noticeably different for those subjects who dropped out early as compared to those that completed later courses (see Figures 1–3). Specifically, the cisplatin/etoposide "completers'" QOL measurements did not deteriorate as quickly as those who dropped out early. Therefore, this missing data pattern was concluded to be nonignorable and subjects were divided into two groups, those who completed no more than the second course and those who completed at least one course beyond the second course. For simplicity, we will refer to the former group as "Dropouts" and the latter group as "Completers" throughout this analysis.

In the Physical Functioning subscale, there was no statistically significant difference between the Taxol/cisplatin arm and the cisplatin/etoposide arm either in Dropouts ($p = 0.50$) or Completers ($p = 0.84$). There was also no statistically significant difference between the Taxol/cisplatin/G-CSF arm and the cisplatin/etoposide arm for either Dropouts ($p = 0.29$) or Completers ($p = 0.61$).

In the Functional Well Being subscale, there was no statistically significant dif-

ference between the Taxol/cisplatin arm and the cisplatin/etoposide arm either in Dropouts ($p = 0.73$) or Completers ($p = 0.49$). There was also no statistically significant difference between the Taxol/cisplatin/G-CSF arm and the cisplatin/etoposide arm for either Dropouts ($p = 0.97$) or Completers ($p = 0.37$).

Finally, in the Lung Cancer Specific Symptoms subscale, there was no statistically significant difference between the Taxol/cisplatin arm and the cisplatin/etoposide arm either in Dropouts ($p = 0.026$) or Completers ($p = 0.021$). There was also no statistically significant difference between the Taxol/cisplatin/G-CSF arm and the cisplatin/etoposide arm for either Dropouts ($p = 0.016$) or Completers ($p = 0.15$). Recall that the nominal Type I error for these comparisons is 0.0125 (not 0.05). These results are consistent with the sponsor's results obtained with the Wei-Lachin test.

The results of the analysis by this reviewer confirms the findings of Fairclough (1997). Namely, there seems to be non-ignorable missing data mechanism in this QOL data set. Therefore, the results of any QOL analysis on these data should be interpreted with caution. This reviewer's analyses did not detect any statistically significant differences between treatment arms for the three QOL endpoints under consideration.

3.3 Summary and Conclusions for CA139-165

Summary

Overall, there seemed to be little evidence as to any increased efficacy of the Taxol arms versus the cisplatin/etoposide arm. Although the Taxol arms generally yielded favorable results in a number of endpoints, the results were not generally statistically significant.

The sponsor performed a "pooled Taxol arm" analysis although it was not prespecified in the protocol. That is, there were two distinct doses of Taxol-containing regimens in this study, each with different doses. The sponsor pooled patients in these arms in order to compare this combined arm with cisplatin/etoposide. The median survival was 7.4 months (95% 6.5-8.6 months) for the cisplatin/etoposide group and 9.7 months (95% 8.8-10.6 months) for the pooled taxol group (stratified log-rank $p = 0.049$). The sponsor concluded that there was a statistically significant difference. However, the nominal Type I error rate must be less than 0.0125 given the fact that this comparison is made in addition to the three comparisons that were

pre-specified in the protocol (see Section 3.1). Therefore, one must conclude that there is no statistically significant difference between survival in the pooled Taxol group and the cisplatin/etoposide group.

The primary endpoint for this study was overall survival. There was no statistically significant difference found between Taxol arms and the cisplatin/etoposide arm with respect to survival ($p = 0.125$ for Taxol/cisplatin vs. cisplatin/etoposide; $p = 0.079$ for Taxol/cisplatin/G-CSF vs. cisplatin/etoposide). The median survival was 7.4 months (95% CI 6.5-8.6 months) for cisplatin/etoposide, 9.3 months (95% CI 8.0-10.4 months) for Taxol/cisplatin, and 10.0 months (95% CI 9.8-11.7 months) for Taxol/cisplatin/G-CSF. Using a stratified Cox model to compare hazard ratios of the Taxol arms versus the cisplatin/etoposide arm resulted in no statistically significant difference ($p = 0.094$) for Taxol/cisplatin/G-CSF vs. cisplatin/etoposide or for Taxol/cisplatin vs. cisplatin/etoposide ($p = 0.169$). This was also confirmed in a Cox regression on survival that included the following covariates: gender, LDH, prior radiotherapy, and histology. LDH as a covariate was a statistically significant contributor to the model, whereas treatment was not (Taxol/cisplatin vs. cisplatin/etoposide $p = 0.169$; Taxol/cisplatin/G-CSF vs. cisplatin/etoposide $p = 0.094$).

The objective response rates were 23% (95% CI 18-30%) for Taxol/cisplatin, 25% (95% CI 20-32%) for Taxol/cisplatin/G-CSF, and 12% (95% CI 8-17%) for cisplatin/etoposide in an intent-to-treat analysis. There was a statistically significant difference between Taxol/cisplatin and the cisplatin/etoposide arm (CMH stratified $p = 0.001$), and a statistically significant difference between Taxol/cisplatin/G-CSF and the cisplatin/etoposide arm (CMH stratified $p < 0.001$). There were no statistically significant differences with respect to response duration for either the Taxol/cisplatin vs. cisplatin/etoposide comparison (stratified log-rank $p = 0.681$) or Taxol/cisplatin/G-CSF vs. cisplatin/etoposide comparison (stratified log-rank $p = 0.227$).

The median time to progression for Taxol/cisplatin was 4.3 months (95% CI 3.3-5.1 months). For Taxol/cisplatin/G-CSF, it was 4.9 months (95% CI 4.0-5.8 months), and for cisplatin/etoposide, it was 2.7 months (95% CI 2.2-3.2 months), There was a statistically significant difference between the Taxol/cisplatin/G-CSF arm and the cisplatin/etoposide arm with respect to time to progression using this reviewer's analysis ($p = 0.007$). This was also confirmed in a Cox regression that included the following covariates: gender, LDH, prior radiotherapy, and histology ($p = 0.005$). There was a no statistically significant difference in time to progression between the Taxol/cisplatin arm and the cisplatin/etoposide arm based on this reviewer's

analysis ($p = 0.091$). A Cox regression including the above covariates resulted in a p-value of 0.060 for comparing Taxol/cisplatin and cisplatin/etoposide. LDH was a statistically significant covariate in these Cox regressions.

Although three QOL analyses were performed, no analysis showed any statistically significant change in QOL between treatment arms over time. The sponsor incorrectly reported a statistically significant improvement of Taxol/cisplatin over cisplatin/etoposide in Lung Cancer Symptoms; the p-value for Lung Cancer Symptoms was 0.027, but the comparison-wise error rate was 0.0125, not 0.05.

The sponsor's QOL results are, in general, difficult to interpret in this reviewer's opinion because the issues of intra-subject correlation and dropout were not addressed. The Wei-Lachin test, which was used to test for differences in QOL in this study, assumes that dropout is completely random; this is not always the case in cancer trials. There was also no Type I error adjustment for the multiplicity of subscales that were considered in the QOL analysis.

Two longitudinal approaches, which addressed these issues, resulted in showing no statistically significant differences between Taxol arms and cisplatin/etoposide with respect to QOL endpoints.

Conclusions

Although the high-dose Taxol arm (Taxol/cisplatin/G-CSF) approached statistically significant improved efficacy compared to cisplatin/etoposide on a number of endpoints, it was only found to be statistically significant on time to progression and overall objective response. However, there was more toxicity and adverse events on this arm; the Medical Review explores this in more detail.

There was a statistically significant difference between Taxol/cisplatin and cisplatin/etoposide only with respect to objective response.

Patient benefit is questionable to this reviewer since there are marked increases in toxicity and no substantial efficacy advantages with either Taxol regimen. The Medical Reviewer will include a detailed analysis of the safety data.