



U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Pharmacoepidemiology and Statistical Science
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number: 21-677 / N000

Drug Name: Alimta (Pemetrexed, LY231514)

Indication(s): Locally advanced or metastatic non-small cell lung cancer

Applicant: Eli Lilly and Company

Date(s): Submission Date: November 3, 2003
ODAC Date: July 27, 2004
Date of Review: June 29, 2004
PDUFA Date: September 3, 2004

Review Priority: Standard

Biometrics Division: Division of Biometrics I (HFD-710)

Statistical Reviewer: Yong-Cheng Wang, Ph.D. (HFD-710)

Concurring Reviewers: Rajeshwari Sridhara, Ph.D. (HFD-710)
Kooros Mahjoob, Ph.D. (HFD-710)

Medical Division: Division of Oncology Drug Products (HFD-150)

Clinical Team: Martin Cohen, M.D. & John Johnson, M.D. (HFD-150)

Project Manager: Ms. Patricia Garvey

Keywords: Post-hoc study objective, Superiority hypothesis, Non-inferiority hypothesis, Fixed margin, Fraction retention, Active control, Multiple comparison, Treatment crossover

Table of Contents

1. EXECUTIVE SUMMARY.....	3
1.1 CONCLUSIONS AND RECOMMENDATIONS	3
1.2 BRIEF OVERVIEW OF CLINICAL STUDIES	3
1.3 STATISTICAL ISSUES AND FINDINGS	4
1.3.1 Major Statistical Issues.....	4
1.3.2 Statistical Findings.....	6
2. INTRODUCTION	9
2.1 OVERVIEW	9
2.1.1 Background.....	9
2.1.2 History of Drug Development	10
2.1.3 Specific Studies Reviewed.....	10
2.2 DATA SOURCES	11
3. STATISTICAL EVALUATION	12
3.1 EVALUATION OF EFFICACY	12
3.1.1 Study Design and Endpoints	12
3.1.2 Patient Dispositions, Demographic and Baseline Characteristics.....	14
3.1.3 Statistical Methodologies	18
3.1.4 Sponsor’s Results	21
3.1.5 Reviewer’s Results.....	27
3.1.6 Reviewer’s Conclusion and Comments	29
3.2 EVALUATION OF SAFETY	29
4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS	30
4.1 GENDER, RACE AND AGE.....	30
4.1.1 Gender	30
4.1.2 Race.....	30
4.1.3 Age.....	31
4.2 OTHER SPECIAL/SUBGROUP POPULATIONS	32
5. SUMMARY AND CONCLUSIONS.....	32
5.1 STATISTICAL ISSUES AND COLLECTIVE EVIDENCE.....	32
5.2 CONCLUSIONS AND RECOMMENDATIONS	34
APPENDIX 1. EXPLORATORY ANALYSES FOR RT POPULATION	35
APPENDIX 2. STATISTICAL PRINCIPLES AND ICH GUIDELINES.....	36
APPENDIX 3. STATISTICAL ISSUES OF ACTIVE CONTROL NON-INFERIORITY TRIALS	37
APPENDIX 4. REFERENCES	40
SIGNATURES/DISTRIBUTION LIST PAGE.....	41

1. EXECUTIVE SUMMARY

1.1 Conclusions and Recommendations

In this statistical reviewer's opinion, the data and results of the single, randomized, open-label, multi-center phase III study H3E-MC-JMEI comparing alimta (pemetrexed, LY231514) to active control docetaxel in patients with locally advanced or metastatic (Stage IIIA, IIIB or IV) non-small cell lung cancer (NSCLC) does not support the applicant's efficacy claim of alimta. This study failed to demonstrate superior efficacy of alimta over docetaxel ($p=0.9300$) for the primary endpoint of overall survival. Furthermore, this study also failed to demonstrate non-inferiority of alimta compared to docetaxel based on a protocol-defined fixed non-inferiority margin (hazard ratio of alimta over docetaxel < 1.11) ($p=0.2558$). The applicant's claim of efficacy ($p=0.047$) was based on 50% retention non-inferiority hypothesis testing and an arbitrary estimate of control effect size which were not pre-specified in the protocol. Furthermore, this p-value was not adjusted for multiple tests and was not significant ($p=0.6395$) when more commonly used estimate of control effect was used. In the presence of crossover of treatment from alimta to docetaxel, it is also difficult to interpret demonstration of non-inferiority. There is no substantial evidence based on this single trial to support the applicant's claim of non-inferiority efficacy of alimta compared to docetaxel.

1.2 Brief Overview of Clinical Studies

This application consists of report of results from registration Study H3E-MC-JMEI (JMEI) in the treatment of patients with NSCLC, supportive data from the single-agent Phase 2 Studies H3E-MC-JMBR (JMBR), H3E-MC-JMAL (JMAL), H3E-MC-JMAN (JMAN), H3E-MC-JMAY (JMAY), H3E-MC-JMBZ (JMBZ), and H3E-MC-JMEK (JMEK) in the treatment of patients with NSCLC.

Study JMBR is the main supporting Phase 2 study of single-agent alimta (500 mg/m^2 every 3 weeks) in patients with NSCLC whose disease was refractory to prior chemotherapy. Two additional Phase 2 studies, JMAL and JMAN, examined the tumor efficacy of single-agent alimta (600 mg/m^2 every 3 weeks) in patients with chemotherapy-naive NSCLC. Three other Phase 2 studies JMAY, JMBZ, and JMEK examined the efficacy of alimta in combination with platinum in first line treatment of patients with NSCLC.

The study selected for this statistical review is Study JMEI which was an international, randomized, Phase 3, active-controlled, open-label, multi-center study to compare alimta with docetaxel in patients with locally advanced or metastatic (Stage IIIA, IIIB or IV) NSCLC who had received prior chemotherapy. A total of 571 patients had been randomly assigned to one of two treatment arms. Alimta was given as a 500 mg/m^2 intravenous infusion on Day 1 of a 21-day cycle. Patients on this arm received folic acid supplementation, 350 to 1000 μg , or equivalent, and injections of 1000 μg vitamin B₁₂. Folic acid was taken orally daily beginning approximately 1 to 2 weeks prior to the first dose of alimta and continued daily until 3 weeks after the last dose of alimta. A vitamin B₁₂ injection was given intramuscularly approximately 1 to 2 weeks prior to the first dose of alimta and was repeated approximately every 9 weeks until 3

weeks after the last dose of alimta. Oral dexamethasone, 4 mg twice per day (or equivalent), was given on the day before, the day of, and the day after alimta therapy, unless it was clinically contraindicated.

Docetaxel was given as a 75 mg/m² intravenous infusion on Day 1 of a 21-day cycle. Patients on this arm received oral dexamethasone, 16 mg per day (for example, 8 mg twice daily), or with an equivalent regimen for 3 days starting the day before docetaxel administration, unless clinical contraindications existed. Patients on the docetaxel treatment arm did not receive folic acid or vitamin B₁₂ supplementation.

1.3 Statistical Issues and Findings

1.3.1 Major Statistical Issues

1. Study failed to demonstrate superiority efficacy per the original protocol objective.
2. Study failed to demonstrate efficacy based on the fixed margin non-inferiority test as defined in the amended protocol.
3. The sponsor claimed the non-inferiority of alimta to docetaxel based on the 50% retention of control (docetaxel) effect non-inferiority testing. However, the sponsor's fraction retention non-inferiority analysis was based on a questionable estimate of control effect which was the middle value of 95% CI of log-hazard ratio of docetaxel to best support care (BSC). Based on FDA analysis the study failed to demonstrate efficacy based on the percent retention of control effect non-inferiority testing. Furthermore, this hypothesis testing approach was a post-hoc addition in the statistical analysis plan after the study was completed and just before the data was locked in this open-label study. Based on the guidance International Conference on Harmonisation (ICH)-E9: Statistical Principles for Clinical Trials, the analysis based on non-inferiority testing using percent retention approach can only be considered as exploratory since this was not pre-specified in the protocol (Appendix 2).
4. Study JMEI was designed to allow subjects to receive the post-study chemotherapy after progression. In alimta group, there were a total of 90 subjects (31.8%; 90/283) who received post-study docetaxel therapy, 42 subjects (14.8%, 42/283) who received other post-study chemotherapy and 151 subjects (53.4%; 151/283) who did not receive any post-study chemotherapy. In docetaxel group, there were a total of 11 subjects (3.8%; 11/288) who received post-study docetaxel therapy, 96 subjects (33.3%, 96/288) who received other post-study chemotherapy and 181 subjects (62.8%; 181/288) who did not receive any post-study chemotherapy. The overall survival results are therefore confounded by the cross-over, and any conclusion based on non-inferiority testing could potentially be biased and uninterpretable.
5. Multiple statistical tests (a superiority test and two non-inferiority tests) for the primary efficacy endpoint have been included in this NDA submission. The two non-inferiority (fixed margin and fraction retention of control effect) hypotheses are not nested within each other. Therefore, the overall significance level after the first non-inferiority (fixed margin) test is not maintained in the second non-inferiority (fraction retention of control effect) test. No multiplicity adjustment has been made in the NDA submission.

6. The control (docetaxel) treatment effect was estimated based on a single, small, randomized trial comparing docetaxel to BSC. The hazard ratio of docetaxel (75 mg/m²) over BSC was 0.56 (95% CI: 0.35 to 0.88) (docetaxel label). The reliability of the estimated control effect is questionable because of single small historical trial. (ICH-E10: Choice of Control Group and Related Issues in Clinical Trials (Appendix 2).)
7. The NDA submission defined a 50% margin for the non-inferiority hypothesis of fraction retention. This was not pre-specified in the protocol. ICH-E10 guidelines states that: “The determination of the margin in a non-inferiority trial is based on both statistical reasoning and clinical judgment, should reflect uncertainties in the evidence on which the choice is based, and should be suitably conservative.”
8. There are two fundamental assumptions in the fraction retention non-inferiority analysis: the control treatment should be truly effective and the control effect has not changed over time (constancy assumption). However, these two assumptions can not be verified since the estimation of control effect is based on a single, small, randomized trial. Inter-trial variability is not included in estimating the active control effect size and therefore it is difficult to determine if the estimated effect is true and reliable.
9. This statistical reviewer has three major concerns regarding the analysis and interpretation submitted in this NDA. (1) The standard statistical comparisons can not be employed in this NDA and p-values are not interpretable based on the post-hoc definition of non-inferiority hypothesis of fraction retention. (2) The p-value presented in the NDA submission were based the sponsor’s estimate of control effect (hazard ratio of docetaxel over BSC = 0.59). The sponsor explained their estimation which was the middle value of 95% CI of log-hazard ratio of control relative to BSC. However, though the estimated log-hazard ratio is proved to be asymptotically normally distributed, the sponsor’s estimate of active control effect may not be appropriate because the selected historical trial is too small (104 patients). (3) Since the active control effect is estimated based on only one small historical trial, the point estimate of hazard ratio may not be appropriate to establish the control effect. To minimize the risk in the overestimation of control effect, a 95% lower confidence limit is generally used as an estimate of the control effect. These results suggest that the p-values from non-inferiority test results are not interpretable.
10. The sponsor claimed that alimta retained 102% of docetaxel’s clinical benefit. This is only a point estimate of fraction retention based on the geometric definition. Since there was only one small historical trial selected for the non-inferiority analysis, the variation would be very large. Therefore, this point estimate is for reference only.
11. The sponsor calculated a 95% CI (52% to 157%) for the fraction retention and claimed that the fraction retention test was based on this CI and there was no multiplicity adjustment needed. This is misleading. By Rothmann’s method, we can only get a conditional CI for the fraction retention, which is not the exact CI for fraction retention. Also, the fraction retention non-inferiority test is totally different test than the fixed margin non-inferiority test.
12. The sponsor claimed that alimta provided a significant survival advantage over BSC (hazard ratio = 0.55; p = 0.019). This is also misleading and not a valid comparison. Since there is no BSC at the current trial (Study JMEI), we can not directly compare alimta to BSC.
13. None of the secondary efficacy analysis demonstrated superior treatment effect of alimta compared to docetaxel.

1.3.2 Statistical Findings

Confirmatory Analyses (Superiority and Fixed Margin Non-inferiority) for the Primary Endpoint

Overall survival was the primary efficacy endpoint of Study JME1. Two statistical tests for the primary endpoint were defined in the protocol amendment: (1) Test for superiority of alimta relative to docetaxel (H_{01} : HR ≥ 1), and (2) Test for non-inferiority based on a protocol-defined fixed margin (H_{02} : HR ≥ 1.11). Since these two tests were pre-specified in the protocol, the analyses based on these two tests are presented below.

Table 1 summarizes the results of the superiority test and fixed margin non-inferiority test of the primary endpoint for ITT population. It failed to reach the significance level 0.05 in superiority test ($p=0.9300$; log-rank) and fixed margin non-inferiority test ($p=0.2558$).

Table 1: Confirmatory Analyses^a of Primary Endpoint: Overall Survival – ITT Population

	Sponsor Analysis		FDA Analysis	
	Alimta (N = 283)	Docetaxel (N = 288)	Alimta (N = 283)	Docetaxel (N = 288)
Events	206	203	206	203
Survival time (months)				
Median	8.3	7.9	8.3	7.9
(95% CI)	(7.0, 9.4)	(6.3, 9.2)	(7.0, 9.4)	(6.3, 9.2)
Superiority test				
p-value of log-rank test ^b	Not reported		0.9300	
p-value of Wilcoxon test ^b	Not reported		0.5944	
Non-inferiority fixed margin test				
p-value of NI fixed margin test ^b	0.226		0.2558	
Hazard ratio ^c	0.99		0.992	
95% CI for hazard ratio ^c	(0.82, 1.20)		(0.817, 1.204)	

^a Superiority and fixed margin non-inferiority analyses as defined in the protocol.

^b P-value is based on the test results for the two treatment groups.

^c Hazard ratio is based on Cox proportional-hazard model with the treatment as single independent variable.

Exploratory Analyses (Fraction Retention Non-inferiority) for the Primary Endpoint

The NDA submission also included a third statistical test for the primary endpoint: Test for non-inferiority based on a percentage of the docetaxel benefit retained by alimta (H_{03} : $\delta \leq 50\%$), where δ is called *fraction retention*. In this trial, it is the percentage of the control (docetaxel) effect retained by alimta. Since this test was not pre-specified in the protocol, the analyses based on this test are considered as exploratory.

In general when only one small historical trial is used to estimate the control effect, use of a point estimate inflates type I error and the lower 95% confidence limit (LCL) is used as the estimate of control effect. However, the sponsor used an arbitrary point estimate in the estimation of the control effect. We report the results of fraction retention non-inferiority tests with two different estimates of control effect in Table 2. It failed to reach the significance level 0.05 in the 50% fraction retention non-inferiority test ($p=0.0525$ and 0.6395 based on the point estimate and 95% LCL of control effect, respectively).

Table 2: Exploratory Analyses^a of Primary Endpoint: Overall Survival – ITT Population

	Sponsor Analysis		FDA Analysis	
	Alimta (N = 283)	Docetaxel (N = 288)	Alimta (N = 283)	Docetaxel (N = 288)
Events	206	203	206	203
Fraction retention non-inferiority test based on the point estimate of control effect (HR(P/C) = 1.7857)				
Non-inferiority fraction retention test				
Estimate of control effect		0.59 ^d		0.56 ^e
NI p-value for testing 50% retention ^b		0.047		0.0525
95% conditional CI of estimated percent of efficacy retained by alimta ^c		(52%, 157%)		(48.56%, 158.97%)
Fraction retention non-inferiority test based on the 95% LCL of control effect (HR(P/C) = 1.1364)				
Non-inferiority fraction retention test				
NI p-value for testing 50% retention ^b		Not reported		0.6395
95% conditional CI of estimated percent of efficacy retained by alimta ^c		Not reported		Not available

^a Fraction retention non-inferiority analyses which were not pre-specified in the protocol.

^b P-value is based on the test results for the two treatment groups by Rothmann's method for a 50% retention.

^c 95% conditional CI is based on the fixed control effect as the estimate by the historical data.

^d The sponsor's estimate based on middle point of 95% CI of log-HR (BSC vs. docetaxel) from historical trial for ITT population.

^e Point estimate of HR in the historical trial for ITT population, published in docetaxel (taxotere) label.

Sensitivity Analysis to Evaluate the Effect of Treatment Confounding due to Crossover

Study JMEI was designed to allow subjects to receive the post-study chemotherapy after progression. In alimta group, there were a total of 90 subjects (31.8%; 90/283) who received post-study docetaxel therapy, 42 subjects (14.8%, 42/283) who received other post-study chemo therapy and 151 subjects (53.4%; 151/283) who did not receive any post-study chemotherapy. In docetaxel group, there were a total of 11 subjects (3.8%; 11/288) who received post-study docetaxel therapy, 96 subjects (33.3%, 96/288) who received other post-study chemotherapy and 181 subjects (62.8%; 181/288) who did not receive any post-study chemotherapy.

To evaluate the effect of treatment crossover, a sensitivity analysis was conducted and the results are summarized in Table 3. For the subgroup of patients who did not receive post-study chemotherapy, the median survival times were 5.8 months (95% CI: 4.5-7.4) and 4.9 months (95% CI: 4.1-6.2) for alimta and docetaxel groups, respectively. The median survival times for the subgroup of patients who received post-study docetaxel therapy were 9.5 months (95% CI: 8.4-10.2) and 10.1 months (95% CI: 7.9-19.5) for alimta and docetaxel groups, respectively. The median survival times for the subgroup of patients who received other post-study chemotherapy were 10.6 months (95% CI: 7.8-14.1) and 11.2 months (95% CI: 9.3-13.9) for alimta and docetaxel groups, respectively.

Table 3: Sensitivity Analysis of Treatment Crossover for Primary Endpoint – FDA Analysis

	ITT Population	
	Alimta (N = 283)	Docetaxel (N = 288)
No post-study chemotherapy		
Number of patients	151	181
Events	114	137
Median survival (months) (95% CI)	5.8 (4.5, 7.4)	4.9 (4.1, 6.2)
Post-study docetaxel therapy		
Number of patients	90	11
Events	67	7
Median survival (months) (95% CI)	9.5 (8.4, 10.2)	10.1 (7.9, 19.5)
Other post-study chemotherapy		
Number of patients	42	96
Events	25	59
Median survival (months) (95% CI)	10.6 (7.8, 14.1)	11.2 (9.3, 13.9)

2. INTRODUCTION

2.1 Overview

2.1.1 Background

Lung cancer is one of the most common malignancies that continues to rise in incidence; one million new cases and over 900,000 lung cancer-related deaths are reported each year worldwide. It is the leading cause of cancer death in men and the third leading cause in women. An estimated 164,000 new cases were diagnosed in the United States in 2000, accounting for approximately 13% of all cancer diagnoses and 28% of all US cancer deaths. Almost 80% of lung cancers are classified as NSCLC, with 65% to 75% of cases presenting as locally advanced (Stage III) or metastatic disease (Stage IV).

Patients diagnosed with Stage III NSCLC generally receive chemotherapy as part of standard multimodality treatment, whereas Stage IV patients typically receive chemotherapy alone as first-line therapy. Historically, NSCLC has not responded well to second-line chemotherapy, and, until recently, no drug had earned regulatory approval in the second-line setting. Single-agent therapy with vindesine, epirubicin, etoposide, or cisplatin showed response rates of $\leq 10\%$.

Docetaxel (taxotere) was approved by the agency in December 1999 in the United States for use in patients with Stage III or IV NSCLC as post-platinum second-line therapy, based on two randomized Phase 3 trials. European Commission approval followed in January 2000. The first trial compared docetaxel 100 mg/m² with best supportive care (BSC) (Shepherd *et al.* 2000). Five deaths in the first 49 enrolled patients led to a docetaxel dose reduction to 75 mg/m². The most common Grade 3 or 4 toxicity was neutropenia (76%). Seven percent of the patients had a partial response in the docetaxel arm, with a median survival of 7.5 months compared with a median survival of 4.6 months for the patients in the BSC arm (p=0.010, based on log-rank test).

The second trial compared docetaxel with navelbine or ifosfamide. A total of 373 patients were randomly assigned to receive docetaxel 100mg/m² or 75 mg/m² (median survival=5.7 months) compared with a control regimen of navelbine or ifosfamide (median survival=5.6 months). The overall response rate was 10.8% with 100 mg/m² docetaxel and 6.7% with 75 mg/m² docetaxel. These response rates were each significantly higher than treatment with navelbine or ifosfamide (0.8%). The 1-year survival rate in the docetaxel arm was significantly better at 32% compared with 19% in the ifosfamide/vinorelbine arm. However, the overall survival was not significantly different among the groups.

Based on these data, the recommended dose of docetaxel for NSCLC patients was 75 mg/m² intravenously every 21 days, preceded by premedication with oral corticosteroids, such as dexamethasone.

Alimta is a novel pyrrolopyrimidine-based antifolate cytotoxic drug jointly discovered by the sponsor, Eli Lilly, and Princeton University. In vitro studies have shown that alimta and its intracellularly polyglutamated metabolites are highly cytotoxic against human leukemia cells (IC₅₀ = 15 nM). The cytotoxicity of alimta and its metabolites is attributed to their ability to

strongly inhibit several key folate-dependent enzymes involved in nucleic acid biosynthesis. These enzymes include thymidylate synthase (TS), dihydrofolate reductase (DHFR), glycinamide ribonucleotide formyltransferase (GARFT), and aminoimidazole carboxamide ribonucleotide formyltransferase (AICARFT). End-product reversal experiments with human leukemia, colorectal, and other cancer cell lines showed that the cytotoxicity of alimta was only partially reversed by thymidine. Effective reversal required both thymidine and hypoxanthine, suggesting that alimta inhibited both pyrimidine and purine biosynthetic pathways. Studies have also shown that cell lines that overexpress TS or that are resistant to raltitrexed, a specific inhibitor of TS, remained partially sensitive to alimta. These data have led to the hypothesis that alimta may have enhanced antitumor activity compared with other antifolates.

Alimta has been approved for the patients with Malignant Pleural Mesothelioma.

2.1.2 History of Drug Development

The sponsor conducted several Phase 2 clinical studies in NSCLC in which alimta was evaluated as single-agent first line, in combination with cisplatin first line, and as single-agent second line treatment. The initially recommended dose of alimta for Phase 2 trials was 600 mg/m² on Day 1 every 21 days. However, toxicities observed in a Phase 2 colorectal study led to a decrease in the alimta dose to 500 mg/m² on Day 1 every 21 days. As this study was ongoing, the initial analyses using a multiple logistic regression model were able to quantify the relative risk of developing toxicities with alimta and generated a validated clinical hypothesis on ways to improve the safety profile of alimta. The levels of pretreatment total plasma homocysteine and methylmalonic acid significantly predicted Grade 4 neutropenia, Grade 4 thrombocytopenia, Grade 3/4 diarrhea, and Grade 3/4 mucositis. Thus, it was postulated that reducing homocysteine levels with folic acid and vitamins B₁₂ supplementation would reduce severe toxicities. Further prospective trials with vitamin supplementation demonstrated that alimta safety profile was improved without affecting the efficacy.

Based on the results from five phase 2 studies, a randomized Phase 3 trial was initiated in NSCLC patients. The main objectives of the current Phase 3 study were to compare the overall efficacy and toxicity profiles of alimta and docetaxel in Stage III and IV NSCLC patients in a second-line setting.

2.1.3 Specific Studies Reviewed

The sponsor has submitted a New Drug Application (NDA 21-677) for standard approval of alimta. This application consists of report of results from registration Study JMEI in the treatment of patients with NSCLC, supportive data from single-agent Phase 2 Studies JM BR, JM AL, JM AN, JM AY, JM BZ, and JM EK in the treatment of patients with NSCLC.

Study JM BR is the main supporting Phase 2 study of single-agent alimta (500 mg/m² every 3 weeks) in patients with NSCLC whose disease was refractory to prior chemotherapy. The tumor response rate of 8.9% observed in this study is consistent with the response rates obtained on

both study arms of JMEI and with the response rates obtained in the Phase 3 trials of 75 mg/m² docetaxel as second-line treatment of NSCLC. The overall median survival of 5.7 months and time to progressive disease of 2.0 months seen in JMBR were also consistent with docetaxel literature results.

Two additional Phase 2 studies, JMAL and JMAN, examined the tumor efficacy of single-agent alimta (600 mg/m² every 3 weeks) in patients with chemotherapy-naive NSCLC. After the first 3 patients enrolled onto JMAN, the protocol was amended to reduce the planned dose of alimta to 500 mg/m². Tumor response rates among patients considered qualified for efficacy analysis were 18% in JMAL and 23% in JMAN, median survival times were 8.4 months and 9.2 months, and times to progressive disease were 4.5 months and 3.8 months, respectively.

Three other Phase 2 studies JMAY, JMBZ, and JMEK examined the efficacy of alimta in combination with platinum in first line treatment of patients with NSCLC. In Studies JMAY and JMBZ, alimta 500 mg/m² plus cisplatin 75 mg/m² were administered every 3 weeks. The overall response rates for patients evaluable for efficacy were 36.1% and 44.8%, respectively. The median survival in these two studies was 10.9 months and 8.9 months, respectively. In the randomized Phase 2 study JMEK, patients were randomized to alimta plus carboplatin or alimta plus oxaliplatin. The overall response rate in the alimta plus carboplatin arm was 31.6% and in the alimta plus oxaliplatin arm was 26.8%. The corresponding median survival was 9.9 months and 9.3 months, respectively.

The study selected for this statistical review is Study JMEI which was an international, randomized, Phase 3, controlled, open-label, multi-center study to compare alimta with docetaxel in patients with locally advanced or metastatic (Stage IIIA, IIIB or IV) NSCLC who had received prior chemotherapy.

2.2 Data Sources

Data used for review is from the electronic submission received on November 3, 2003. The efficacy analysis data were submitted by the sponsor on December 23, 2003. All data sets analyzed are electronic documents and are located in the Electronic Document Room (EDR) of CDER of FDA under the Letter Date "3-NOV-2003" and "23-DEC-2003". The data sets analyzed in this NDA review are located in the folders of CRT\datasets. The major data sets for the efficacy analyses are "SURVPOP", "PATDEMOG", and "PATSUMM" which defined the survival time, responses, events, time to events, and demographic variables.

3. STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

3.1.1 Study Design and Endpoints

Study JMEI was an international, randomized, Phase 3, controlled, open-label, multi-center study to compare alimta with docetaxel in patients with locally advanced or metastatic (Stage IIIA, IIIB or IV) NSCLC who had received prior chemotherapy. A total of 520 NSCLC patients with measurable or evaluable disease were to be enrolled in this study. However, due to ethical reasons and rapid patient enrollment across all investigative sites, all patients who signed the informed consent document (ICD) were allowed to participate in the trial. As a result, when the study was closed to enrollment, a total of 571 patients had been randomly assigned to one of two treatment arms. Alimta was given as a 500 mg/m² intravenous infusion on Day 1 of a 21-day cycle. Patients on this arm received folic acid supplementation, 350 to 1000 µg, or equivalent, and injections of 1000 µg vitamin B₁₂. Folic acid was taken orally daily beginning approximately 1 to 2 weeks prior to the first dose of alimta and continued daily until 3 weeks after the last dose of alimta. A vitamin B₁₂ injection was given intramuscularly approximately 1 to 2 weeks prior to the first dose of alimta and was repeated approximately every 9 weeks until 3 weeks after the last dose of alimta. Oral dexamethasone, 4 mg twice per day (or equivalent), was given on the day before, the day of, and the day after alimta therapy, unless it was clinically contraindicated.

Docetaxel was given as a 75 mg/m² intravenous infusion on Day 1 of a 21-day cycle. Patients on this treatment arm received oral dexamethasone, 16 mg per day (for example, 8 mg twice daily), or with an equivalent regimen for 3 days starting the day before docetaxel administration, unless clinical contra-indications existed. Patients on the docetaxel treatment arm did not receive the folic acid or vitamin B₁₂ supplementation as described above.

After the initial dose, modifications of alimta or docetaxel doses were allowed, based on patient toxicity. After patients discontinued from study therapy, they proceeded to the post-study follow-up phase of the study. Patients were followed up until death or until lost to follow-up.

The primary objective of this study was to compare overall survival following treatment with alimta versus docetaxel in patients with locally advanced or metastatic (Stage IIIA, IIIB or IV) NSCLC who had been previously treated with chemotherapy.

The secondary objectives of the study were as follows:

- to characterize and compare the quantitative and qualitative toxicities of alimta and docetaxel in this patient population
- to compare the objective tumor response rate of both therapies
- to compare time -to-event efficacy variables of both therapies, including:
 - 1) duration of response
 - 2) time to objective tumor response
 - 3) time to treatment failure

- 4) time to documented disease progression
- 5) progression-free survival.
- to compare changes in the average symptom burden index between the alimta and docetaxel arms by using the Lung Cancer Symptom Scale (LCSS).

The baseline stratification factors included the following:

- ECOG performance status (Low [2] or High [0 or 1])
- prior platinum-containing chemotherapy (Yes or No)
- prior paclitaxel-containing chemotherapy (Yes or No)
- number of prior chemotherapy (1 or 2)
- time since last chemotherapy (<3 months or ≥3 months)
- best response to last prior chemotherapy (CR/PR or SD or PD or unknown)
- disease stage (IIIA, IIIB, or IV).

This study was designed to enroll at least 520 patients, randomly and evenly assigned to treatment between the two treatment arms of alimta or docetaxel. This sample size was chosen based on consideration of the primary comparison (superiority hypothesis) of overall survival between treatment arms to detect alimta superior to docetaxel, with 85% power and at 0.05 the level of significance.

The study protocol design was based on the assumption that in overall survival, the hazard ratio (HR) of alimta to docetaxel is approximately constant over the period of observation.

Superiority of alimta in overall survival was defined by $HR < 1.00$. Non-inferiority of alimta in overall survival was defined by $HR < 1.11$ (protocol amendment (b) dated 03 August 2001 after enrollment 167 number of patients, the first patient enrolled on 20 March 2001). Hazard ratio was estimated from the study data by using the Cox proportional hazards model with therapy arm as the only cofactor.

The primary analyses were performed on the intention to treat (ITT) basis. The ITT population was defined as all patients randomly assigned to a treatment arm whether or not they received study drug. Thus, the ITT population in this study consists of 283 patients in the alimta arm and 288 patients in the docetaxel arm.

No interim analyses were planned for this study.

Reviewer's Comments:

- 1) The original protocol defined the study objective as a superiority to determine whether alimta is more effective than docetaxel. A protocol amendment redefined the study objective as a fixed margin non-inferiority test if the superiority hypothesis is failed. No changes were made regarding the design or sample size when this non-inferiority hypothesis was added.
- 2) The applicant has added post-hoc study objective to test for non-inferiority based on fraction retention if both the superiority test and the pre-defined non-inferiority test based

on fixed margin failed (final statistical analysis plan (SAP) dated 24 January 2003, last patient enrolled on 06 February 2002, treatment completed on 13 November 2002, data locked on 30 January 2003). This additional analysis can only be considered as exploratory.

- 3) Study JMEI was designed to allow subjects to receive the post-study docetaxel or other chemotherapy after progression. However, the NDA submission did not clearly describe the treatment crossover in the study design. This treatment crossover could potentially confound the efficacy results and pose difficulty in interpreting non-inferiority claim.

3.1.2 Patient Dispositions, Demographic and Baseline Characteristics

This was a multi-center trial that entered 698 patients at 135 investigational sites in 23 countries. Of these, 571 (81.8%) patients were randomly assigned (enrolled) to either the alimta arm or the docetaxel arm. The following figure shows the patient population disposition. Of the 698 patients entered, 283 patients were randomly assigned to the alimta arm, and 288 patients were randomly assigned to the docetaxel arm. A total of 114 patients did not meet the protocol inclusion criteria, and 13 patients could not be randomized because of unspecified reasons.

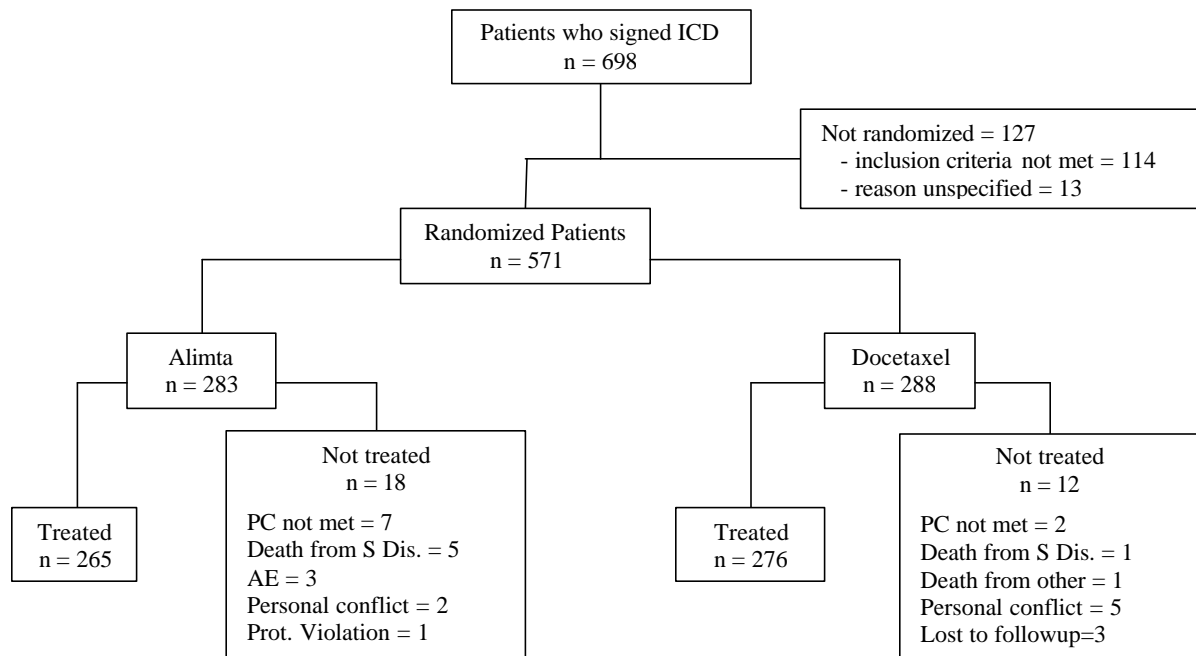


Table 4 presents the key demographic characteristics for the intent-to-treat (ITT) population. Of the 571 subjects enrolled, 72% (411) were male and 28% (160) were female. The proportion of females was somewhat higher in the alimta treatment arm (31.4%; 89/283) than in the docetaxel arm (24.7%; 71/288) for the ITT population. The two randomized treatment groups were similar with respect to origin and age: over 70% (403/571) of the subjects were of Caucasian origin and the mean age was about 58.3 years (range 22.3 to 87.4 years). At baseline, the demographics of the alimta and docetaxel groups were comparable.

Table 4: Subject Demographic Characteristics

Demographic	ITT Population	
	Alimta (N = 283)	Docetaxel (N = 288)
Gender (n%)^a		
Male	194 (68.6)	217 (75.3)
Female	89 (31.4)	71 (24.7)
Race (n%)^a		
African Descent	8 (2.8)	8 (2.8)
Western Asian	20 (7.1)	23 (8.0)
Caucasian	203 (71.7)	200 (69.4)
East/Southeast A	44 (15.6)	49 (17.0)
Hispanic	4 (1.4)	6 (2.1)
Other	4 (1.4)	2 (0.7)
Age (n%)^b		
N	283	288
Mean ± SD	59.0 ± 10.5	58.6 ± 9.5
Range	22.3 – 81.2	29.0 – 87.4
Age Group (n%)^b		
≤ 64	199 (70.3)	214 (74.3)
65 – 74	69 (24.4)	68 (23.6)
≥ 75	15 (5.3)	6 (2.1)

^a The sponsor's analyses verified by the statistical reviewer.

^b FDA's analyses.

Table 5 summarizes the number of patients included in the various stratification factors that were incorporated into the randomization process for the ITT population. These strata were chosen as potential confounders to survival and other study outcomes as suggested in the literature.

Results indicated that the two treatment arms were balanced with respect to most of the prognostic factors. A majority of patients on both arms had good performance status. The most common histological diagnosis among patients was adenocarcinoma, followed by squamous cell carcinoma of the lung; 91.2% of the patients received prior platinum and 26.8% prior taxanes. Thirty-three patients (5.8%) received two regimens of prior chemotherapy.

Table 5: Baseline Stratification Factors Used for Randomization^a

Demographic	ITT Population	
	Alimta (N = 283)	Docetaxel (N = 288)
Performance Status (n%)	264	274
ECOG PS 0	52 (19.7)	48 (17.5)
ECOG PS 1	182 (68.9)	192 (70.1)
ECOG PS 2	30 (11.4)	34 (12.4)
Histological Subtype (n%)	283	288
Adenocarcinoma	154 (54.4)	142 (49.3)
Bronchoalveolar	4 (1.4)	1 (0.3)
Squamous	78 (27.6)	93 (32.3)
Other	47 (16.6)	52 (18.1)
Homocysteine (n%)	283	286
Low (< 12 umol/L)	202 (71.4)	197 (68.9)
High (≥ 12 umol/L)	81 (28.6)	89 (31.1)
Stage of Disease (n%)	283	288
Stage IIIA	14 (4.9)	13 (4.5)
Stage IIIB	57 (20.1)	60 (20.8)
Stage IV	212 (74.9)	215 (74.7)
Prior Chemotherapy (by Regimen#) (n%)	283	288
1 Regimen	270 (95.4)	268 (93.1)
2 Regimen	13 (4.6)	20 (6.9)
Prior Platinum (n%)	283	288
Had No Prior Platinum	21 (7.4)	29 (10.1)
Had Prior Platinum	262 (92.6)	259 (89.9)
Prior Taxane (n%)	283	288
Had No Prior Taxane	210 (74.2)	208 (72.2)
Had Prior Taxane	73 (25.8)	80 (27.8)
Best Response to Chemotherapy (n%)	283	288
Complete Response	12 (4.2)	4 (1.4)
Partial Response	89 (31.4)	101 (35.1)
Stable Disease	106 (37.5)	93 (32.3)
Progressive Disease	67 (23.7)	73 (25.3)
Unknown or Not Done	4 (1.4)	11 (3.8)
Not Evaluable	5 (1.8)	6 (2.1)
Time Since Last Chemotherapy (n%)	278	285
< 3 mos since last chemo	140 (50.4)	137 (48.1)
> 3 mos since last chemo	138 (49.6)	148 (51.9)

^a The sponsor's analyses verified by the statistical reviewer.

Table 6 presents summary of the primary histological diagnoses for the ITT patients by treatment arm. All baseline disease characteristics were balanced between the two treatment arms.

Table 6: Disease Characteristics Histologic Diagnoses^a

Histologic Diagnosis	ITT Population	
	Alimta (N = 283)	Docetaxel (N = 288)
	n (%)	n (%)
NSCLC	22	14
Lung, NSCLC	2 (0.7)	2 (0.7)
NSCLC	9 (3.2)	6 (2.1)
Undifferentiated carcinoma	1 (0.4)	1 (0.3)
Poor differentiated NSCLC	10 (3.5)	5 (1.7)
Adenocarcinoma	154	142
Lung, adenocarcinoma	151 (53.4)	141 (49.0)
Adeno NSC type	2 (0.7)	1 (0.3)
Mucinous adenoca	1 (0.4)	0
Squamous cell carcinoma	78 (27.6)	93
Lung, squamous	75 (26.5)	91 (31.6)
Squamous cell carcinoma	2 (0.7)	2 (0.7)
Squamous cell lung	1 (0.4)	0
Large cell carcinoma	18	29
Lung, large cell	18 (6.4)	29 (10.1)
Other	11	10
Adenoid cyst cancer	0	1 (0.3)
Epidemoid squamous	0	1 (0.3)
Bronchoalveolar adeno carcinoma	1 (0.4)	0
Bronchoalveolar carcinoma	1 (0.4)	1 (0.3)
Lung, adeno-squamous	4 (1.4)	5 (.17)
Lung, bronchoalveolar	2 (0.7)	0
Other unspecified	1 (0.4)	0
Poor differentiated	2 (0.7)	1 (0.3)
Sar. Pleural mesothelioma	0	1 (0.3)

^aThe sponsor's analyses verified by the statistical reviewer.

NSC = non-small cell; NSCLC = non-small cell lung cancer.

Table 7 summarizes the reasons for study discontinuations for the ITT patients by treatment arm. The most common reason for discontinuation in both populations for both treatment arms was lack of efficacy due to progressive disease. More patients (55.5%; 157/283) on the alimta arm discontinued from the study because of lack of efficacy due to progressive disease compared with those (46.9%; 135/288) on the docetaxel arm.

Table 7: Reasons for Discontinuations^a

Reason	ITT Population	
	Alimta (N = 283) n (%)	Docetaxel (N = 288) n (%)
Adverse event	21 (7.4)	25 (8.7)
Clinical relapse ^b	4 (1.4)	1 (0.3)
Death (other causes)	8 (2.8)	11 (3.8)
Death from study disease	14 (4.9)	20 (6.9)
Death related to study drug toxicity	1 (0.4)	3 (1.0)
Lack of efficacy, patient and/or physician perception	25 (8.8)	25 (8.7)
Lack of efficacy, progressive disease	157 (55.5)	135 (46.9)
Lost to follow-up	1 (0.4)	4 (1.4)
Patient has completed therapy	14 (4.9)	21 (7.3)
Patients continuing	1 (0.4)	0
Personal conflict or patient decision	12 (4.2)	18 (6.3)
Protocol entry criteria not met	8 (2.8)	3 (1.0)
Protocol violation	4 (1.4)	4 (1.4)
Satisfactory response patient and/or physician perception	13 (4.6)	18 (6.2)

^aThe sponsor's analyses verified by the statistical reviewer.

^bProgressive disease after complete response or partial response.

3.1.3 Statistical Methodologies

The NDA application described the statistical analysis plan (SAP) as follows, where the superiority test and non-inferiority test based on a protocol-defined fixed margin were pre-specified in the protocol and the non-inferiority test based on at least 50% retention of docetaxel effect was not pre-specified in the protocol.

The primary endpoint of Study JMEI was overall survival time. The primary analysis of Study JMEI was the estimation of the overall survival hazard ratio (HR) between alimta and 75 mg/m² docetaxel. Time-to-event analyses were performed on the observed distributions of overall survival time. Overall survival time was defined as the time from the date of randomization to date of death due to any cause. Overall survival time was censored at the date of the last follow-up visit for patients who were still alive. The primary analysis was the comparison of overall survival between the two study treatment arms in the ITT population. The Cox proportional hazards model (with study treatment arm as the only cofactor) was used to calculate a 95% confidence interval (CI) for this overall survival HR of alimta to docetaxel. Medians for survival time were estimated by regimen using the Kaplan-Meier (K-M) method. Overall survival rates at 3, 6, 9 and 12 months were estimated using the K-M method and compared between regimens based on normal approximations for the differences between rates.

The following three primary tests of statistical hypotheses were performed, based on this 95% CI for the survival HR:

(1) Test for superiority of alimta relative to docetaxel (H_{01} : HR \geq 1).

(2) Test for non-inferiority based on a protocol-defined fixed margin (H_{02} : HR ≥ 1.11). A non-inferiority fixed margin was defined in the JMEI study protocol by a survival HR (alimta over docetaxel) of less than 1.11. In other words, if the upper bound of the 95% CI for this HR was less than 1.11, a statistically significant non-inferiority would be demonstrated. Non-inferiority of alimta to docetaxel using fixed margin would be achieved if the overall survival in the alimta arm is $\leq 10\%$ worse than that observed in the docetaxel arm. This would translate to an upper bound of the 95% CI < 1.11 for the HR of alimta over docetaxel.

(3) Test for non-inferiority based on at least 50% retention of docetaxel effect. Percentage of the docetaxel benefit retained by alimta (H_{03} : $\delta \leq 50\%$), where δ is the percentage of the docetaxel effect retained by alimta which is called *fraction retention*.

To estimate the control treatment (docetaxel) effect, a randomized phase 3 trial (Shepherd *et al.* 2000) was selected where 104 patients were randomly assigned to receive either 75 mg/m² docetaxel or corresponding best supportive care (BSC). The HR of docetaxel over BSC was estimated to be 0.56 (95% CI: 0.35 to 0.88).

Percentage of docetaxel effect over BSC, which is retained by alimta, was calculated based on the following method:

$$\delta = 1 - [\log \text{HR (alimta over docetaxel)} \div \log \text{HR (BSC over docetaxel)}]$$

The 95% CI of this percentage of benefit was calculated using Rothman's Z* statistic (Rothmann *et al.* 2003).

Other time-to-event analyses were performed on the observed distributions of progression-free survival, time to treatment failure (TTTF), and time to documented disease progression (TTPD). The analysis was the comparison of time-to-event variables between the two study treatment arms in the ITT population.

Progression-free survival time was defined as the time from the date of randomization to the first date of documented disease progression or death due to any cause. Progression-free survival time was censored at the date of the last follow-up visit for patients who were still alive and who had not progressed.

Time to treatment failure was defined as the time from the date of randomization to the date of the first of the following events: discontinuation of study therapy, progression of disease, or death due to any cause. Time to treatment failure was censored at the date of the last follow-up visit for patients who did not discontinue, who were still alive, and who did not have disease progression.

Time to documented disease progression was defined as the time from the date of randomization to the first date of documented disease progression. TTPD was censored at the date of death for patients who have not had documented disease progression. For patients who were still alive at

the time of analysis and who did not have documented disease progression, TTPD was censored at the date of the last follow-up visit.

For each of the time-to-event endpoints, the Cox proportional hazards model (with therapy arm as the only cofactor) was used to estimate the respective true HR of alimta to docetaxel. Medians for each of the time-to-event endpoints were estimated by regimen using the K-M method.

Time-to-event variables at 3, 6, etc. months were estimated using the K-M method and compared between regimens based on normal approximations for the differences between rates.

A tumor responder was defined as any patient exhibiting a best study response of complete response (CR) or partial response (PR) (based on CT, MRI, or plain x-ray, and/or palpation) or partial response in non-measurable disease (PRNM).

Response rates, time to objective tumor response, duration of response, and duration of clinical benefit were compared between the treatment arms on the population of CR/PR/PRNM patients.

No interim analysis was performed for the study.

No planned multiplicity adjustments were made to any of the analyses.

Reviewer's Comments:

- 1) The NDA submission stated that “the protocol for this study was approved on 07 November 2000 and was amended on 27 November 2000. The final SAP was approved on 24 January 2003. The reporting database was validated and locked on 30 January 2003.” However, the NDA submission did not state who approved the final SAP. Per agency’s request, the sponsor explained that the final SAP was internally approved. Therefore, post-hoc definition of study objective to add fraction retention non-inferiority test if the superiority and fixed margin non-inferiority hypotheses both failed, was not presented to FDA prior to NDA submission.
- 2) The post-hoc definition of statistical hypothesis, statistical analysis plan and data analyses, may result in a biased efficacy analysis and conclusion. This hypothesis testing is exploratory in nature.
- 3) The NDA submission stated that “no planned multiplicity adjustments were made to any of the analyses.” However, since the two non-inferiority (fixed margin and fraction retention of control effect) hypotheses are not nested within each other, the overall significance level after the first non-inferiority (fixed margin) test is not maintained in the second non-inferiority (fraction retention of control effect) test. No multiplicity adjustments were made.
- 4) The control (docetaxel) effect was estimated based on a single, small, randomized trial comparing docetaxel to best supportive care (BSC). The hazard ratio of docetaxel (75 mg/m²) over BSC was 0.56 (95% CI: 0.35 to 0.88) (docetaxel label). The reliability of the estimated control effect is questionable because of single small historical trial.

- 5) The NDA submission defined a 50% margin for the non-inferiority hypothesis of fraction retention. This was not pre-specified in the protocol. The method of estimation of control effect size was not pre-specified and was not agreed upon by FDA.
- 6) The NDA submission quoted that “in Shepherd’s trial, where 104 patients were randomly assigned to receive either 75 mg/m² docetaxel or corresponding BSC, the HR of docetaxel over BSC was estimated to be 0.56 (95% CI: 0.35 to 0.88).” However, the sponsor did not use 0.56 as the estimate of the control (docetaxel) effect, which was published in the docetaxel label. A different value 0.59 was used in the sponsor’s SAS program of non-inferiority test of fraction retention. Per agency’s request, the sponsor explained their calculation based on the middle value of 95% CI of log-hazard ratio of BSC relative to docetaxel. However, though the estimated log-hazard ratio is proved to be asymptotically normally distributed, the sponsor’s estimate of control effect may not be appropriate because the selected historical trial is too small (104 patients). Furthermore, since the control effect is estimated based on only one small historical trial, the point estimate of hazard ratio may not be appropriate to establish the control effect. To minimize the risk in the overestimation of control effect, a 95% lower confidence limit is generally used as an estimate of the control effect.
- 7) The NDA submission stated the fraction retention non-inferiority test as “setting the percentage of historical benefit at 50% and maintaining an approximate one-sided 2.5% type I error, an upper 95% CI bound of < 1.21 for the HR of alimta over docetaxel is required to establish the non-inferiority of alimta”. This is an incorrect interpretation of the fraction retention non-inferiority analysis. The fraction retention non-inferiority hypothesis is $H_{03}: \delta \leq 50\%$, where δ is a ratio of two hazard ratios (treatment vs. active control and control vs. to placebo) defined as above. In other words, a fixed margin non-inferiority analysis will test a $HR \geq$ a fixed margin (constant). A fraction retention non-inferiority analysis will test a ratio of two hazard ratios \leq a fixed percentage. They are two totally different hypotheses. The methods of statistical inferences for the two hypotheses are totally different. The historical data are treated as constants in a fixed margin non-inferiority test but treated as random variables in a fraction retention non-inferiority test. Therefore, a multiplicity adjustment is definitely needed for the two tests.
- 8) The sponsor also performed additional analyses for the randomized and treated (RT) population. Since there was no information for the RT population from the historical trial, the sponsor used the estimated control effect from ITT population for their non-inferiority analyses of RT population. Therefore, these analyses are for exploratory purposes only. (reported in the appendix of this statistical review.)

3.1.4 Sponsor’s Results

Overall Survival

The primary endpoint for Study JMEI was defined as the overall survival. Table 8 summarizes the results of the overall survival time and two non-inferiority tests for the primary endpoint. For ITT population, the two median survival times were 8.3 (95% CI: 7.0 – 9.4) months and 7.9 (95% CI: 6.3 - 9.2) months for the alimta and docetaxel groups, respectively. The superiority

analysis was not reported in the NDA submission. The study also failed to reach significance level 0.05 in the fixed margin non-inferiority test (p=0.226). Since the sponsor used a different estimate for the control effect, the p-value of 50% fraction retention non-inferiority test is not interpretable. The survival curves for the overall survival by K-M estimate are presented in Figure 1.

Table 8: Primary Efficacy Endpoint: Overall Survival ^a		
	ITT Population	
	Alimta (N = 283)	Docetaxel (N = 288)
Events	206	203
Survival time (months)		
Median	8.3	7.9
(95% CI)	(7.0, 9.4)	(6.3, 9.2)
Non-inferiority fixed margin test		
p-value of NI fixed margin test ^b		0.226
Hazard ratio ^c		0.99
95% CI for hazard ratio ^c		(0.82, 1.20)
Non-inferiority fraction retention test		
NI p-value for testing 50% retention ^d		0.047
95% conditional CI of estimated percent of efficacy retained by alimta ^e		(52%, 157%)

^a The sponsor's analyses.

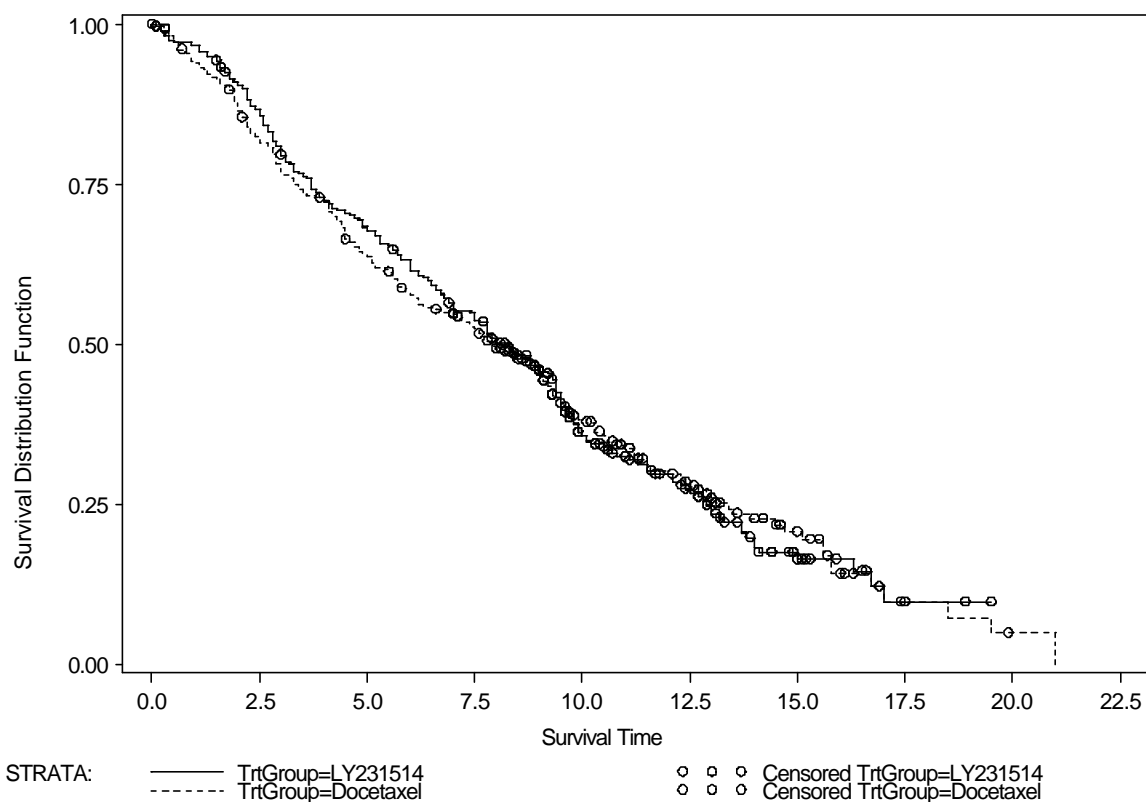
^b P-value is based on the test results for the two treatment groups, not adjusted for multiplicity.

^c Hazard ratio is based on Cox proportional-hazard model with the treatment as single independent variable.

^d P-value is based on the test results for the two treatment groups by Rothmann's method for a 50% retention.

^e Point estimate and 95% conditional CI are based on the fixed control effect estimated by the docetaxel trial

Figure 1 . Kaplan – Meier Curve of Survival Time
Population = ITT



Progression-free Survival

Table 9 summarizes the results of the statistical analysis for progression-free survival time. For ITT population, the two median survival times were 2.9 (95% CI: 2.4 – 3.1) months and 2.9 (95% CI: 2.7 – 3.4) months for the alimta and docetaxel groups, respectively. P-value based on the log-rank test was 0.756 and HR of alimta to docetaxel was 0.973 (95% CI: 0.82 – 1.16). The survival curves for the progression-free survival by K-M estimate are presented in Figure 2.

Table 9: Secondary Efficacy Endpoint: Progression-free Survival^a

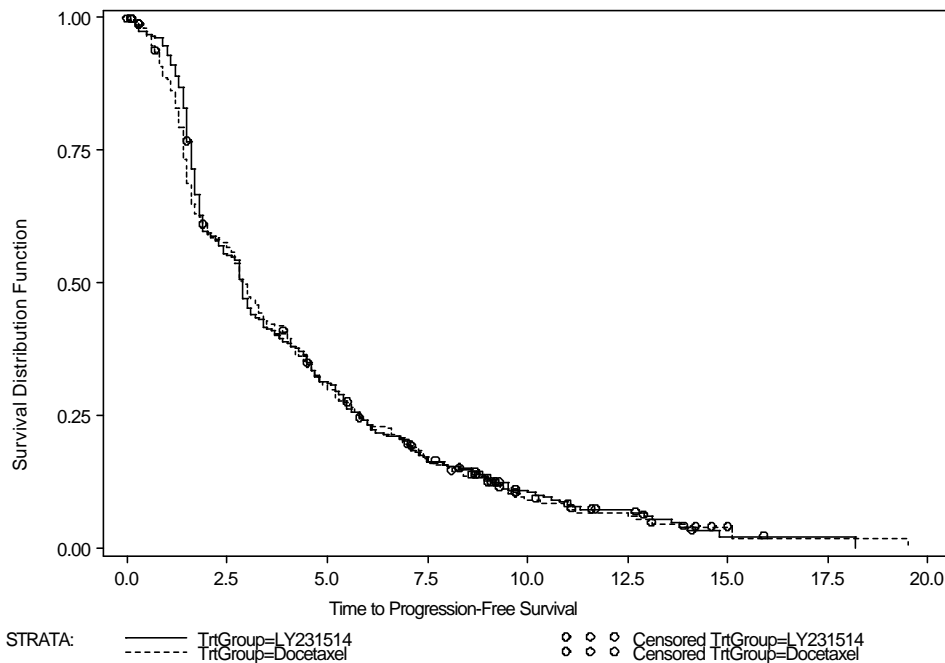
	ITT Population	
	Alimta (N = 283)	Docetaxel (N = 288)
Events	265	258
Survival time (months)		
Median	2.9	2.9
(95% CI)	(2.4, 3.1)	(2.7, 3.4)
Superiority test		
p-value of log-rank test ^b		0.756
p-value of Wilcoxon test ^b		0.419
Hazard ratio ^c		0.973
95% CI for hazard ratio ^c		(0.82, 1.16)

^a The sponsor's analyses.

^b P-value is based on the test results for the two treatment groups.

^c Hazard ratio is based on Cox proportional-hazard model with the treatment as single independent variable.

**Figure 2 . Kaplan-Meier Curve of Time to Progression-Free Survival
Population = ITT**



Time to Progressive Disease

Table 10 summarizes the results of the statistical analysis for time to progressive disease (TTPD). In the ITT population, the TTPD for the alimta arm was similar to the docetaxel arm (median time 3.4 months versus 3.5 months). P-value based on the log-rank test was 0.721 and the HR of alimta to docetaxel was 0.97 with the 95% HR CI of 0.80 to 1.17.

Table 10: Secondary Efficacy Endpoint: Time to Progressive Disease^a

	ITT Population	
	Alimta (N = 283)	Docetaxel (N = 288)
Survival time (months)		
Minimum	0.5	0.3
25 th percentile	1.7	1.5
Median	3.4	3.5
75 th percentile	7.0	7.3
Maximum	18.2	19.5
Superiority test		
p-value of log-rank test ^b		0.721
Hazard ratio ^c		0.97
95% CI for hazard ratio ^c		(0.80, 1.17)

^a The sponsor's analyses.

^b P-value is based on the test results for the two treatment groups.

^c Hazard ratio is based on Cox proportional-hazard model with the treatment as single independent variable.

Time to Treatment Failure

Table 11 summarizes the results of the statistical analysis for time to treatment failure. For ITT population, the two median survival times were 2.3 (95% CI: 1.8 – 2.8) months and 2.1 (95% CI: 1.7 – 2.8) months for the alimta and docetaxel groups, respectively. P-value based on the log-rank test was 0.041 and HR of alimta to docetaxel was 0.842 (95% CI: 0.71 – 0.995). The survival curves for the time to treatment failure by K-M estimate are presented in Figure 3.

Table 11: Secondary Efficacy Endpoint: Time to Treatment Failure^a

	ITT Population	
	Alimta (N = 283)	Docetaxel (N = 288)
Events	278	283
Survival time (months)		
Median (95% CI)	2.3 (1.8, 2.8)	2.1 (1.7, 2.8)
Superiority test		
p-value of log-rank test ^b		0.041
p-value of Wilcoxon test ^b		0.064
Hazard ratio ^c		0.842
95% CI for hazard ratio ^c		(0.71, 0.995)

^a The sponsor's analyses.

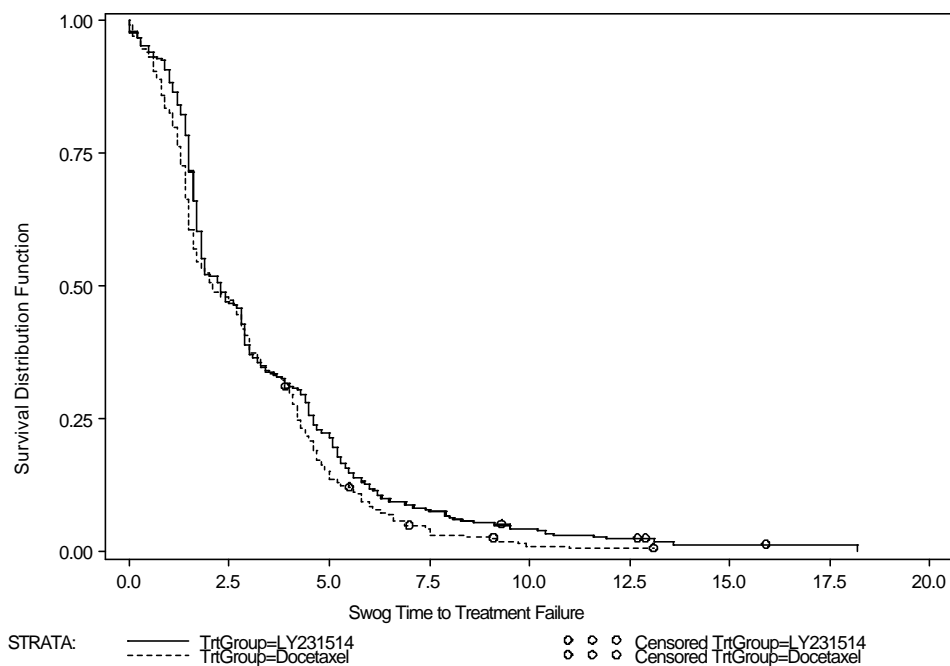
^b P-value is based on the test results for the two treatment groups.

^c Hazard ratio is based on Cox proportional-hazard model with the treatment as single independent variable.

Reviewer's Comments:

This endpoint is generally not acceptable as it includes toxicity events.

Figure 3 . Kaplan–Meier Curve of Swog Time to Treatment Failure
Population = ITT



Best Tumor Response

Table 12 summarizes the results of the statistical analysis for the investigator-determined best tumor response for the population which were qualified for tumor response (QR) analysis. The response rate for the alimta and docetaxel groups were 24 (9.09%; 95% CI: 5.9 – 13.2) and 24 (8.76%; 95% CI: 5.7 – 12.8), respectively. P-value based on Chi-square test was 0.893. The number of patients with the best response of complete response (CR), partial response (PR), partial response in non-measurable disease (PRNM), progressive disease (PD), stable disease (SD), or unknown (U) were similar between the two treatment arms.

Table 12: Secondary Efficacy Endpoint: Best Tumor Response^a

	QR Population	
	Alimta (N = 264)	Docetaxel (N = 274)
Response (%)		
Complete response	1 (0.38)	0
Partial response	20 (7.58)	24 (8.76)
Partial response in non-measurable disease	3 (1.14)	0
Progressive disease	97 (36.74)	93 (33.94)
Stable disease	121 (45.83)	127 (46.35)
Unknown	22 (8.33)	30 (10.95)
Response rate analysis		
Response rate (CR+PR+PRNM) (%)	24 (9.09)	24 (8.76)
95% CI for response rate	(5.9, 13.2)	(5.7, 12.8)
p-value of Chi-square test ^b	0.893	

^a The sponsor’s analyses.

^b P-value is based on the test results for the two treatment groups.

3.1.5 Reviewer’s Results

Confirmatory Analyses (Superiority and Fixed Margin Non-inferiority) for the Primary Endpoint

Overall survival was the primary efficacy endpoint of Study JMEI. Two statistical tests for the primary endpoint were defined in the protocol amendment: (1) Test for superiority of alimta relative to docetaxel (H_{01} : $HR \geq 1$), and (2) Test for non-inferiority based on a protocol-defined fixed margin (H_{02} : $HR \geq 1.11$). Since these two tests were pre-specified in the protocol, the analyses based on these two tests are confirmatory in the application.

Table 13 summarizes the results of the superiority test and fixed margin non-inferiority test of the primary endpoint for ITT population. It failed to reach the significance level 0.05 in superiority test ($p=0.9300$; log-rank) and fixed margin non-inferiority test ($p=0.2558$).

Table 13: Confirmatory Analyses^a of Primary Endpoint: Overall Survival – ITT Population

	Sponsor Analysis		FDA Analysis	
	Alimta (N = 283)	Docetaxel (N = 288)	Alimta (N = 283)	Docetaxel (N = 288)
Events	206	203	206	203
Survival time (months)				
Median	8.3	7.9	8.3	7.9
(95% CI)	(7.0, 9.4)	(6.3, 9.2)	(7.0, 9.4)	(6.3, 9.2)
Superiority test				
p-value of log-rank test ^b	Not reported		0.9300	
p-value of Wilcoxon test ^b	Not reported		0.5944	
Non-inferiority fixed margin test				
p-value of NI fixed margin test ^b	0.226		0.2558	
Hazard ratio ^c	0.99		0.992	
95% CI for hazard ratio ^c	(0.82, 1.20)		(0.817, 1.204)	

^a Superiority and fixed margin non-inferiority analyses as defined in the protocol.

^b P-value is based on the test results for the two treatment groups.

^c Hazard ratio is based on Cox proportional-hazard model with the treatment as single independent variable.

Exploratory Analyses (Fraction Retention Non-inferiority) for the Primary Endpoint

The NDA submission also included a third statistical test for the primary endpoint: Test for non-inferiority based on a percentage of the docetaxel benefit retained by alimta (H_{03} : $\delta \leq 50\%$), where δ is called *fraction retention*. In this trial, it is the percentage of the control (docetaxel) efficacy retained by alimta. Since this test was not pre-specified in the protocol, the analyses based on this test are considered as exploratory.

When only one small historical trial is used to estimate the control effect, use of a point estimate inflates type I error and the 95% lower confidence limit (LCL) is generally used as the estimate of the control effect. However, the sponsor used an arbitrary point estimate in the estimation of the control effect. We report the results of fraction retention non-inferiority tests with two

different estimates of control effect in Table 14. The study failed to reach the significance level 0.05 in the 50% fraction retention non-inferiority test ($p=0.0525$ and 0.6395 based on the point estimate and 95% LCL of control effect, respectively).

Table 14: Exploratory Analyses^a of Primary Endpoint: Overall Survival – ITT Population

	Sponsor Analysis		FDA Analysis	
	Alimta (N = 283)	Docetaxel (N = 288)	Alimta (N = 283)	Docetaxel (N = 288)
Events	206	203	206	203
Fraction retention non-inferiority test based on the point estimate of control effect (HR(P/C) = 1.7857)				
Non-inferiority fraction retention test				
Estimate of control effect		0.59 ^d		0.56 ^e
NI p-value for testing 50% retention ^b		0.047		0.0525
95% conditional CI of estimated percent of efficacy retained by alimta ^c		(52%, 157%)		(48.56%, 158.97%)
Fraction retention non-inferiority test based on the 95% LCL of control effect (HR(P/C) = 1.1364)				
Non-inferiority fraction retention test				
NI p-value for testing 50% retention ^b		Not reported		0.6395
95% conditional CI of estimated percent of efficacy retained by alimta ^c		Not reported		Not available

^a Fraction retention non-inferiority analyses which were not pre-specified in the protocol.

^b P-value is based on the test results for the two treatment groups by Rothmann's method for a 50% retention.

^c 95% conditional CI is based on the fixed control effect as the estimate by the historical data.

^d The sponsor's estimate based on middle point of 95% CI of log-HR (BSC vs. docetaxel) from historical trial for ITT population.

^e Point estimate of HR in the historical trial for ITT population, published in docetaxel (taxotere) label.

Sensitivity Analysis to Evaluate the Effect of Treatment Confounding due to Treatment Crossover

Study JMEI was designed to allow subjects to receive the post-study chemotherapy after progression. In alimta group, there were a total of 90 subjects (31.8%; 90/283) who received post-study docetaxel therapy, 42 subjects (14.8%, 42/283) who received other post-study chemotherapy and 151 subjects (53.4%; 151/283) who did not receive any post-study chemotherapy. In docetaxel group, there were a total of 11 subjects (3.8%; 11/288) who received post-study docetaxel therapy, 96 subjects (33.3%, 96/288) who received other post-study chemotherapy and 181 subjects (62.8%; 181/288) who did not receive any post-study chemotherapy.

To evaluate the effect of treatment crossover, a sensitivity analysis was conducted and the results are summarized in Table 15. For ITT population, the median survival times for the no post-study chemotherapy are 5.8 months (95% CI: 4.5-7.4) and 4.9 months (95% CI: 4.1-6.2) for alimta and docetaxel groups, respectively. The median survival times for post-study docetaxel therapy are 9.5 months (95% CI: 8.4-10.2) and 10.1 months (95% CI: 7.9-19.5) for alimta and docetaxel groups, respectively. The median survival times for other post-study chemotherapy are 10.6 months (95% CI: 7.8-14.1) and 11.2 months (95% CI: 9.3-13.9) for alimta and docetaxel groups, respectively.

Table 15: Sensitivity Analysis of Treatment Crossover for Primary Endpoint – FDA Analysis

	ITT Population	
	Alimta (N = 283)	Docetaxel (N = 288)
No post-study chemotherapy		
Number of patients	151	181
Events	114	137
Median survival (months) (95% CI)	5.8 (4.5, 7.4)	4.9 (4.1, 6.2)
Post-study docetaxel therapy		
Number of patients	90	11
Events	67	7
Median survival (months) (95% CI)	9.5 (8.4, 10.2)	10.1 (7.9, 19.5)
Other post-study chemotherapy		
Number of patients	42	96
Events	25	59
Median survival (months) (95% CI)	10.6 (7.8, 14.1)	11.2 (9.3, 13.9)

3.1.6 Reviewer’s Conclusion and Comments

The pivotal trial H3E-MC-JMEI failed to demonstrate superior efficacy of alimta over docetaxel (p=0.9300) for the primary endpoint: overall survival. Using a closed procedure, this study failed to demonstrate non-inferiority based on a protocol-defined fixed non-inferiority margin (hazard ratio of alimta over docetaxel < 1.11) (p=0.2558). Furthermore, it failed to demonstrate non-inferiority of 50% retention of docetaxel effect by alimta (p=0.0525 and 0.6395 based on the point estimate and 95% LCL of control effect, respectively). Study JMEI also failed to demonstrate superior efficacy of alimta over docetaxel with respect to the progression-free survival (p=0.756), time to progressive disease (p=0.721), and tumor response rate (p=0.893).

Study JMEI was designed to allow subjects to receive the post-study chemotherapy after progression. A sensitivity analysis was conducted. For ITT population, the median survival times for the no post-study chemotherapy are 5.8 months (95% CI: 4.5-7.4) and 4.9 months (95% CI: 4.1-6.2) for alimta and docetaxel groups, respectively. The median survival times for post-study docetaxel therapy are 9.5 months (95% CI: 8.4-10.2) and 10.1 months (95% CI: 7.9-19.5) for alimta and docetaxel groups, respectively. The median survival times for other post-study chemotherapy are 10.6 months (95% CI: 7.8-14.1) and 11.2 months (95% CI: 9.3-13.9) for alimta and docetaxel groups, respectively.

3.2 Evaluation of Safety

No safety evaluation is included in this statistical review.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race and Age

This statistical reviewer conducted the efficacy analysis for gender, race and age subgroups. Since there was no data reported in the NDA submission about the control (docetaxel) effects of gender, race and age subgroups, non-inferiority tests have not been done for these subgroups.

4.1.1 Gender

The efficacy analyses of overall survival, progression-free survival, time to treatment failure, and best tumor response for ITT population by gender are summarized in Table 16.

Table 16: Gender Subgroup Analysis for ITT Population – FDA Analysis

	Female Subgroup		Male Subgroup	
	Alimta (N = 89)	Docetaxel (N = 71)	Alimta (N = 194)	Docetaxel (N = 217)
Overall survival time (months)				
Events (n)	63	40	143	163
Median (95% CI)	8.8 (7.5, 9.9)	11.5 (7.9, 13.4)	7.8 (6.0, 9.4)	6.9 (5.2, 8.7)
Hazard ratio (95% CI) ^a	1.28 (0.86, 1.92)		0.95 (0.76, 1.19)	
p-value of superiority test ^b	0.222		0.629	
Progression-free survival (months)				
Events (n)	84	61	181	197
Median (95% CI)	3.1 (2.7, 4.4)	4.1 (2.9, 4.9)	2.8 (1.9, 3.1)	2.8 (2.1, 3.2)
Hazard ratio (95% CI) ^a	1.20 (0.86, 1.68)		0.91 (0.75, 1.12)	
p-value of superiority test ^b	0.269		0.363	
Time to treatment failure (months)				
Events (n)	87	69	191	214
Median (95% CI)	2.4 (1.8, 3.0)	3.2 (2.1, 4.3)	2.2 (1.8, 2.8)	1.9 (1.6, 2.5)
Hazard ratio (95% CI) ^a	1.04 (0.75, 1.43)		0.78 (0.64, 0.95)	
p-value of superiority test ^b	0.809		0.011	
Best tumor response				
Response rate (n/N)	15.0% (12/80)	7.3% (5/69)	6.5% (12/184)	9.3% (19/205)
p-value of Chi-square test ^c	0.138		0.318	

^a Hazard ratio is based on Cox proportional-hazard model with the treatment as single independent variable.

^b P-value is based on the superiority (log-rank) test results for the two treatment groups.

^c P-value is based on the Chi-square test results for the two treatment groups.

4.1.2 Race

The efficacy analyses of overall survival, progression-free survival, time to treatment failure, and best tumor response for ITT population by race are summarized in Table 17.

Table 17: Race Subgroup Analysis for ITT Population – FDA Analysis

	Caucasian		Others	
	Alimta (N = 203)	Docetaxel (N = 200)	Alimta (N = 80)	Docetaxel (N = 88)
Overall survival time (months)				
Events (n)	148	152	58	51
Median (95% CI)	8.3 (6.7, 9.5)	7.6 (5.6, 9.1)	8.0 (6.8, 9.5)	9.2 (7.2, 12.3)
Hazard ratio (95% CI) ^a	0.91 (0.73, 1.15)		1.27 (0.87, 1.87)	
p-value of superiority test ^b	0.440		0.220	
Progression-free survival (months)				
Events (n)	190	186	75	72
Median (95% CI)	2.8 (2.2, 3.1)	2.8 (2.0, 3.1)	3.1 (2.3, 4.4)	3.9 (2.8, 5.2)
Hazard ratio (95% CI) ^a	0.90 (0.74, 1.11)		1.15 (0.83, 1.59)	
p-value of superiority test ^b	0.318		0.411	
Time to treatment failure (months)				
Events (n)	200	197	78	86
Median (95% CI)	2.2 (1.8, 2.8)	1.9 (1.6, 2.6)	2.3 (1.8, 3.0)	2.8 (2.0, 3.9)
Hazard ratio (95% CI) ^a	0.77 (0.63, 0.94)		1.05 (0.77, 1.43)	
p-value of superiority test ^b	0.009		0.790	
Best tumor response				
Response rate (%)	9.6% (18/187)	6.2% (12/193)	7.8% (6/77)	14.8% (12/81)
p-value of Chi-square test ^c	0.218		0.165	

^a Hazard ratio is based on Cox proportional-hazard model with the treatment as single independent variable.

^b P-value is based on the superiority (log-rank) test results for the two treatment groups.

^c P-value is based on the Chi-square test results for the two treatment groups.

4.1.3 Age

The efficacy analyses of overall survival, progression-free survival, time to treatment failure, and best tumor response for ITT population by age group are summarized in Table 18.

Table 18: Age Subgroup Analysis for ITT Population – FDA Analysis

	< 65 years		≥ 65 years	
	Alimta (N = 199)	Docetaxel (N = 214)	Alimta (N = 84)	Docetaxel (N = 74)
Overall survival time (months)				
Events (n)	146	149	60	54
Median (95% CI)	7.9 (6.8, 9.3)	7.8 (5.7, 9.3)	8.9 (6.5, 10.0)	8.8 (6.2, 10.3)
Hazard ratio (95% CI) ^a	0.95 (0.76, 1.20)		1.15 (0.79, 1.68)	
p-value of superiority test ^b	0.660		0.456	
Progression-free survival (months)				
Events (n)	189	189	76	69
Median (95% CI)	2.9 (2.2, 3.0)	2.9 (2.5, 3.4)	3.1 (2.3, 4.5)	2.9 (2.0, 4.0)
Hazard ratio (95% CI) ^a	1.04 (0.85, 1.27)		0.81 (0.58, 1.12)	
p-value of superiority test ^b	0.705		0.195	
Time to treatment failure (months)				
Events (n)	195	210	83	73
Median (95% CI)	2.3 (1.8, 2.9)	2.4 (1.7, 2.8)	2.3 (1.8, 2.8)	2.0 (1.5, 2.9)
Hazard ratio (95% CI) ^a	0.82 (0.67, 1.002)		0.87 (0.63, 1.19)	
p-value of superiority test ^b	0.052		0.381	
Best tumor response				
Response rate (%)	11.2% (21/188)	9.8% (20/205)	4.0% (3/76)	5.8% (4/69)
p-value of Chi-square test ^c	0.647		0.604	

^a Hazard ratio is based on Cox proportional-hazard model with the treatment as single independent variable.

^b P-value is based on the superiority (log-rank) test results for the two treatment groups.

^c P-value is based on the Chi-square test results for the two treatment groups.

4.2 Other Special/Subgroup Populations

No other special or subgroup analysis is included in this statistical review.

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

1. Study failed to demonstrate superiority efficacy per the original protocol objective.
2. Study failed to demonstrate efficacy based on the fixed margin non-inferiority test as defined in the amended protocol.
3. The sponsor claimed the non-inferiority of alimta to docetaxel based on the 50% retention of control (docetaxel) effect non-inferiority testing. However, the sponsor's fraction retention non-inferiority analysis was based on a questionable estimate of control effect which was the middle value of 95% CI of log-hazard ratio of docetaxel to best support care (BSC). Based on FDA analysis the study failed to demonstrate efficacy based on the percent retention of control effect non-inferiority testing. Furthermore, this hypothesis testing approach was a post-hoc addition in the statistical analysis plan after the study was completed and just before the data was locked in this open-label study. Based on the guidance International Conference on Harmonisation (ICH)-E9: Statistical Principles for Clinical Trials, the analysis based on

non-inferiority testing using percent retention approach can only be considered as exploratory since this was not pre-specified in the protocol (Appendix 2).

4. Study JMEI was designed to allow subjects to receive the post-study chemotherapy after progression. In alimta group, there were a total of 90 subjects (31.8%; 90/283) who received post-study docetaxel therapy, 42 subjects (14.8%, 42/283) who received other post-study chemotherapy and 151 subjects (53.4%; 151/283) who did not receive any post-study chemotherapy. In docetaxel group, there were a total of 11 subjects (3.8%; 11/288) who received post-study docetaxel therapy, 96 subjects (33.3%, 96/288) who received other post-study chemotherapy and 181 subjects (62.8%; 181/288) who did not receive any post-study chemotherapy. The overall survival results are therefore confounded by the cross-over, and any conclusion based on non-inferiority testing could potentially be biased and uninterpretable.
5. Multiple statistical tests (a superiority test and two non-inferiority tests) for the primary efficacy endpoint have been included in this NDA submission. The two non-inferiority (fixed margin and fraction retention of control effect) hypotheses are not nested within each other. Therefore, the overall significance level after the first non-inferiority (fixed margin) test is not maintained in the second non-inferiority (fraction retention of control effect) test. No multiplicity adjustment has been made in the NDA submission.
6. The control (docetaxel) treatment effect was estimated based on a single, small, randomized trial comparing docetaxel to BSC. The hazard ratio of docetaxel (75 mg/m²) over BSC was 0.56 (95% CI: 0.35 to 0.88) (docetaxel label). The reliability of the estimated control effect is questionable because of single small historical trial. (ICH-E10: Choice of Control Group and Related Issues in Clinical Trials (Appendix 2).)
7. The NDA submission defined a 50% margin for the non-inferiority hypothesis of fraction retention. This was not pre-specified in the protocol. ICH-E10 guidelines states that: “The determination of the margin in a non-inferiority trial is based on both statistical reasoning and clinical judgment, should reflect uncertainties in the evidence on which the choice is based, and should be suitably conservative.”
8. There are two fundamental assumptions in the fraction retention non-inferiority analysis: the control treatment should be truly effective and the control effect has not changed over time (constancy assumption). However, these two assumptions can not be verified since the estimation of control effect is based on a single, small, randomized trial. Inter-trial variability is not included in estimating the active control effect size and therefore it is difficult to determine if the estimated effect is true and reliable.
9. This statistical reviewer has three major concerns regarding the analysis and interpretation submitted in this NDA. (1) The standard statistical comparisons can not be employed in this NDA and p-values are not interpretable based on the post-hoc definition of non-inferiority hypothesis of fraction retention. (2) The p-value presented in the NDA submission were based the sponsor’s estimate of control effect (hazard ratio of docetaxel over BSC = 0.59). The sponsor explained their estimation which was the middle value of 95% CI of log-hazard ratio of control relative to BSC. However, though the estimated log-hazard ratio is proved to be asymptotically normally distributed, the sponsor’s estimate of active control effect may not be appropriate because the selected historical trial is too small (104 patients). (3) Since the active control effect is estimated based on only one small historical trial, the point estimate of hazard ratio may not be appropriate to establish the control effect. To minimize the risk in the overestimation of control effect, a 95% lower confidence limit is generally

used as an estimate of the control effect. These results suggest that the p-values from non-inferiority test results are not interpretable.

10. The sponsor claimed that alimta retained 102% of docetaxel's clinical benefit. This is only a point estimate of fraction retention based on the geometric definition. Since there was only one small historical trial selected for the non-inferiority analysis, the variation would be very large. Therefore, this point estimate is for reference only.
11. The sponsor calculated a 95% CI (52% to 157%) for the fraction retention and claimed that the fraction retention test was based on this CI and there was no multiplicity adjustment needed. This is misleading. By Rothmann's method, we can only get a conditional CI for the fraction retention, which is not the exact CI for fraction retention. Also, the fraction retention non-inferiority test is totally different test than the fixed margin non-inferiority test.
12. The sponsor claimed that alimta provided a significant survival advantage over BSC (hazard ratio = 0.55; p = 0.019). This is also misleading and not a valid comparison. Since there is no BSC at the current trial (Study JMEI), we can not directly compare alimta to BSC.
13. None of the secondary efficacy analysis demonstrated superior treatment effect of alimta compared to docetaxel.

5.2 Conclusions and Recommendations

In this statistical reviewer's opinion, the data and results of the single, randomized, open-label, multi-center phase III study H3E-MC-JMEI comparing alimta (pemetrexed, LY231514) to active control docetaxel in patients with locally advanced or metastatic (Stage IIIA, IIIB or IV) non-small cell lung cancer (NSCLC) does not support the applicant's efficacy claim of alimta. This study failed to demonstrate superior efficacy of alimta over docetaxel (p=0.9300) for the primary endpoint of overall survival. Furthermore, this study also failed to demonstrate non-inferiority of alimta compared to docetaxel based on a protocol-defined fixed non-inferiority margin (hazard ratio of alimta over docetaxel < 1.11) (p=0.2558). The applicant's claim of efficacy (p=0.047) was based on 50% retention non-inferiority hypothesis testing and an arbitrary estimate of control effect size which were not pre-specified in the protocol. Furthermore, this p-value was not adjusted for multiple tests and was not significant (p=0.6395) when more commonly used estimate of control effect was used. In the presence of crossover of treatment from alimta to docetaxel, it is also difficult to interpret demonstration of non-inferiority. There is no substantial evidence based on this single trial to support the applicant's claim of non-inferiority efficacy of alimta compared to docetaxel.

APPENDIX 1. EXPLORATORY ANALYSES FOR RT POPULATION

Table 19 summarizes the results of the superiority test and fixed margin non-inferiority test of the primary endpoint for RT population. It failed to reach the significance level 0.05 in superiority test (p=0.7654; log-rank) and fixed margin non-inferiority test (p=0.1879). The results of the fraction retention non-inferiority test are summarized in Table 20.

Table 19: Exploratory Analyses^a of Primary Endpoint: Overall Survival – RT Population

	Sponsor Analysis		FDA Analysis	
	Alimta (N = 265)	Docetaxel (N = 276)	Alimta (N = 265)	Docetaxel (N = 276)
Events	192	198	192	198
Survival time (months)				
Median	8.4	8.0	8.4	8.0
(95% CI)	(7.4, 9.4)	(6.7, 9.2)	(7.4, 9.4)	(6.7, 9.2)
Superiority test				
p-value of log-rank test ^b	Not reported		0.7654	
p-value of Wilcoxon test ^b	Not reported		0.3940	
Non-inferiority fixed margin test				
p-value of NI fixed margin test ^b	0.155		0.1879	
Hazard ratio ^c	0.97		0.971	
95% CI for hazard ratio ^c	(0.80, 1.18)		(0.795, 1.184)	

^a Superiority and fixed margin non-inferiority analyses as defined in the protocol.

^b P-value is based on the test results for the two treatment groups.

^c Hazard ratio is based on Cox proportional-hazard model with the treatment as single independent variable.

Table 20: Exploratory Analyses^a of Primary Endpoint: Overall Survival – RT Population

	Sponsor Analysis		FDA Analysis	
	Alimta (N = 265)	Docetaxel (N = 276)	Alimta (N = 265)	Docetaxel (N = 276)
Events	192	198	192	198
Fraction retention non-inferiority test based on the point estimate of control effect (HR(P/C) = 1.7857)				
Non-inferiority fraction retention test				
Estimate of control effect	0.59 ^d		0.56 ^e	
NI p-value for testing 50% retention ^b	0.036		0.0399	
95% conditional CI of estimated percent of efficacy retained by alimta ^c	(58%, 168%)		(56.12%, 171.48%)	
Fraction retention non-inferiority test based on the 95% LCL of control effect (HR(P/C) = 1.1364)				
Non-inferiority fraction retention test				
NI p-value for testing 50% retention ^b	Not reported		0.5480	
95% conditional CI of estimated percent of efficacy retained by alimta ^c	Not reported		Not available	

^a Fraction retention non-inferiority analyses which were not pre-specified in the protocol.

^b P-value is based on the test results for the two treatment groups by Rothmann's method for a 50% retention.

^c 95% conditional CI is based on the fixed control effect as the estimate by the historical data.

^d The sponsor's estimate based on middle point of 95% CI of log-HR (BSC vs. docetaxel) from historical trial for ITT population.

^e Point estimate of HR in the historical trial for ITT population, published in docetaxel (taxotere) label.

APPENDIX 2. STATISTICAL PRINCIPLES AND ICH GUIDELINES

1. The NDA submission redefined the study objective from the original protocol and protocol amendment. The ICH-E9 guidelines, section A of Considerations for Overall Clinical Development, states that “A confirmatory trial is an adequately controlled trial in which the hypotheses are stated in advance and evaluated. As a rule, confirmatory trials are necessary to provide firm evidence of efficacy or safety. In such trials the key hypothesis of interest follows directly from the trial's primary objective, is always predefined, and is the hypothesis that is subsequently tested when the trial is complete ... Confirmatory trials are intended to provide firm evidence in support of claims; hence adherence to protocols and standard operating procedures is particularly important.”
2. The NDA submission selected only one randomized phase III historical trial (docetaxel label) to establish the control (docetaxel) effect. However, the choice of control group and historical trials are always a critical decision in designing a clinical trial. The ICH-E10 guidelines, section of Introduction, states that “that choice affects the inferences that can be drawn from the trial, the ethical acceptability of the trial, the degree to which bias in conducting and analyzing the study can be minimized, the types of subjects that can be recruited and the pace of recruitment, the kind of endpoints that can be studied, the public and scientific credibility of the results, the acceptability of the results by regulatory authorities, and many other features of the study, its conduct, and its interpretation.”
3. The NDA submission defined a 50% margin for the non-inferiority hypothesis of fraction retention. However, the ICH-E10 guidelines, section E of Introduction, states that “an acceptable non-inferiority margin should be defined, taking into account the historical data and relevant clinical and statistical considerations.” “This margin is the degree of inferiority of the test treatments to the control that the trial will attempt to exclude statistically. If the confidence interval for the difference between the test and control treatments excludes a degree of inferiority of the test treatment as large as, or larger than, the margin, the test treatment can be declared non-inferior; if the confidence interval includes a difference as large as the margin, the test treatment cannot be declared non-inferior.” “The determination of the margin in a non-inferiority trial is based on both statistical reasoning and clinical judgment, **should reflect uncertainties in the evidence on which the choice is based, and should be suitably conservative.** If this is done properly, a finding that the confidence interval for the difference between new drug and the active control excludes a suitably chosen margin provides assurance that the test drug has an effect greater than zero.”

APPENDIX 3. STATISTICAL ISSUES OF ACTIVE CONTROL NON-INFERIORITY TRIALS

The NDA submission included a fraction retention non-inferiority analysis in which 50% of the control effect was expected to be retained by the experimental treatment. Since there are many statistical issues regarding the active control non-inferiority analysis which are under discussion within statistical theory and application, some background on these issues are presented in this appendix.

For a clinical trial involving life-threatening disease, it is considered unethical to use placebo as a control (Temple [1], Temple and Ellenberg [2], Ellenberg and Temple [3], Fleming [4]). In such trial, an available active drug or treatment, or the best current standard of care is usually used as the control. For life-threatening diseases, this kind of active control trials has become an important tool for demonstrating that a new treatment or therapy is effective.

Let T and C denote the new treatment and the active control respectively. Let P denote the placebo, if a placebo were present in the trial. Let C_C and P_C denote the active control and placebo respectively in a non-concurrent (historical) trial. Let HR stand for the hazard ratio.

Traditionally, the effectiveness of a new treatment is demonstrated by showing that it is non-inferior to, or no worse than, the control by a certain pre-specified fixed margin λ_0 (Blackwelder [5]). In statistical terms, if the trial outcome rejects the following null hypothesis H_0 at the desired level of significance, then we may conclude that the new treatment is non-inferior to the control relative to the pre-specified fixed margin λ_0 .

$$H_0: HR(T/C) \geq 1 + \lambda_0 \quad \text{vs.} \quad H_a: HR(T/C) < 1 + \lambda_0,$$

where λ_0 is an arbitrary fixed non-inferiority margin. Although the test of the above null hypothesis H_0 is straight forward, the real question is how one pre-specifies the fixed margin λ_0 .

If the fixed margin λ_0 is chosen arbitrarily, then the trial may run the risk of showing that the new treatment is non-inferior to the control when in fact it is worse than a placebo. For example, if the fixed margin λ_0 is larger than the true control effect, $HR(P_C/C_C) - 1$, then there is a high probability that one may conclude that the new treatment is non-inferior to the control, when in fact it loses all the effect of the control, $HR(P_C/C_C) - 1$, or it could even be worse than a placebo if one were to be present (Chi *et al.* [6]). To minimize such risk, the obvious strategy is assure that the fixed margin λ_0 is less than the control effect $HR(P_C/C_C) - 1$ by setting $\lambda_0 = (1 - d_0)[HR(P_C/C_C) - 1]$, where d_0 is the fraction of the control effect $HR(P_C/C_C) - 1$ that one wishes to retain. Therefore, the margin λ_0 can simply be interpreted as the fractional loss of the control effect that one is willing to tolerate.

With λ_0 so specified, the preceding fixed margin hypothesis becomes,

$$H_0: HR(T/C) \geq 1 + (1 - d_0)[HR(P_C/C_C) - 1] \quad \text{vs.}$$

$$H_a: HR(T/C) < 1 + (1-d_0)[HR(P\check{C}\check{Q}-I)].$$

However, this hypothesis is really not a fixed margin hypothesis, since the margin, λ_0 , depends on the true control effect, $HR(P\check{C}\check{Q}-I$, which is unknown and needs to be estimated. In addition, since in the concurrent trial, there is no placebo, one can not really estimate this control effect. Therefore, to provide an estimate of the control effect, one needs to have some non-concurrent (historical) trials that can provide reliable estimate of the control effect, $HR(P\check{C}\check{Q}-I$, and furthermore, one needs to assume that if a placebo were to be present in the current trial, the true control effect, $HR(P\check{C}\check{Q}-I$, would be maintained in the current trial (constancy assumption). Lastly, one must assume that the current trial has assay sensitivity, that is, the trial is capable of detecting a positive treatment effect if the treatment is truly effective.

Under these various assumptions, one strategy is to estimate the unknown control effect, $HR(P\check{C}\check{Q}-I$, by the lower limit of the 90% or 95% confidence interval of the estimate of the non-concurrent control effect, $HR(P\check{C}\check{Q}-I$, and define λ_0 as half of this control effect estimate. This strategy has been criticized as being too conservative. An alternative strategy is to estimate the unknown control effect, $HR(P\check{C}\check{Q}-I$, by the point estimate of the control effect and define λ_0 as half of this point estimate. However, this strategy is criticized as being too liberal as shown by Rothmann *et al.* [7].

In order to overcome these limitations and realizing that the above hypothesis involves two unknown parameters, Rothmann *et al.* [7] has proposed a method for testing a non-inferiority hypothesis H_0 that is defined in terms of a combination of the two unknown parameters, $HR(T/C)$ and $HR(P\check{C}\check{Q})$ as shown below.

$$\begin{aligned} H_0: [HR(T/C)-1] - (1-d_0)[HR(P\check{C}\check{Q}-1)] &\geq 0 \quad \text{vs.} \\ H_a: [HR(T/C)-1] - (1-d_0)[HR(P\check{C}\check{Q}-1)] &< 0. \end{aligned}$$

Rothmann has argued for the use of the parameters, $\log HR(T/C)$ and $\log HR(P\check{C}\check{Q})$, instead of the parameters, $HR(T/C)$ and $HR(P\check{C}\check{Q})$. But the motivation behind the method is the same.

To test the above linear combination of $HR(T/C)$ and $HR(P\check{C}\check{Q})$, we can define the fraction of the control effect to be retained by the new treatment by

$$d = \frac{HR(P'/C') - HR(T/C)}{HR(P'/C') - 1}$$

and the corresponding fraction retention hypothesis by

$$H_0: d \leq d_0 \quad \text{vs.} \quad H_a: d > d_0$$

In an active control trial, one may wish to show that the new treatment retains at least $100\delta_0\%$ of the control effect, provided it has other clinically meaningful benefit such as better side effects profile, ease of treatment, etc. that is not available with the control. Typically, for non-inferiority

claim, δ_0 can be set at 0.5. If the new treatment were to show a better than 50% retention of the control effect, then the new treatment would have demonstrated clinically meaningful benefit, even though it may not retain all the effect expected of the control. If the new treatment does not have any other clinically meaningful benefit, then δ_0 may need to be set at a higher level. However, it is not clear what fraction of the control effect one should require the new treatment to retain in order to support the claim that the new treatment is non-inferior or equivalent to the control. On the other hand, to show a 0% retention of the control effect is not deemed acceptable, because ethically it is not justified to use such new treatment if it loses all the effect expected of the control.

Under the assumption that $HR(P\check{C}\check{C} - 1) > 0$, the fraction retention hypothesis would be equivalent to the previous linear hypothesis. Rothmann has developed a test statistic Z^* by

$$Z^* = \frac{\hat{HR}(T/C) - (1 - \mathbf{d}_0)\hat{HR}(P'/C')}{s.e.[\hat{HR}(T/C) - (1 - \mathbf{d}_0)\hat{HR}(P'/C')]}$$

and argued that testing the linear hypothesis using the test statistic Z^* is equivalent to testing the fraction retention hypothesis under this assumption.

Thus, under the above assumption, using the test statistic Z^* to test the linear hypothesis would be legitimate, and one may conclude based on the test, whether a new treatment retains the desired fraction of the control effect. The method proposed by Rothmann *et al.* [7] has been applied to two trials in the evaluation of Xeloda for the treatment of patients with colorectal cancer (FDA [8]).

APPENDIX 4. REFERENCES

1. Temple, R. *Problem in interpreting active control equivalence trials*. *Accountability in Research*, 1996; **4**:267--275.
2. Temple, R. and Ellenberg, S. S. *Placebo-controlled trials and active-control trials in the evaluation of new treatments - Part 1: Ethical and scientific issues*. *Annals of Internal Medicine*, 2000; **133**:455--463.
3. Ellenberg, S. S. and Temple, R. *Placebo-controlled trials and active-control trials in the evaluation of new treatments - Part 2: Practical issues and specific cases*. *Annals of Internal Medicine*, 2000; **133**:464--470.
4. Fleming, T. R. *Treatment evaluation in active control studies*. *Cancer Treatment Reports*, 1987; **71**:1061--1064.
5. Blackwelder, W. *Proving the null hypothesis in clinical trials*. *Control. Clin. Trials*, 1982; **3**:345-353.
6. Chi, G. Y. H., Chen, G., Rothmann, M., and Li, N. *Active control trials*. *Encyclopedia of Biopharmaceutical Statistics*, 2003; 9--15.
7. Rothmann, M., Li, N., Chen, G., Chi, G. Y. H., Temple, R., and Tsou, H. H. *Design and analysis of non-inferiority mortality trials in oncology*. *Statistics in Medicine*, 2003; **22**:239--264.
8. FDA. *Medical-Statistical review for Xeloda (NDA 20-896)*. FDA Division of Freedom of Information, Rockville, MD, dated 23 April, 2001.

SIGNATURES/DISTRIBUTION LIST PAGE

Primary Statistical Reviewer: Yong-Cheng Wang, Ph.D.
Date of Review: June 29, 2004

Concurring Reviewers: Rajeshwari Sridhara, Ph.D.
Kooros Mahjoob, Ph.D.

Statistical Acting Team Leader: Rajeshwari Sridhara, Ph.D.

Biometrics Acting Division Director: Kooros Mahjoob, Ph.D.

cc:

HFD-150 / Ms. Patricia Garvey
HFD-150 / Martin Cohen, M.D.
HFD-150 / John Johnson, M.D.
HFD-150 / Richard Pazdur, M.D.
HFD-710 / Yong-Cheng Wang, Ph.D.
HFD-710 / Rajeshwari Sridhara, Ph.D.
HFD-710 / Kooros Mahjoob, Ph.D.
HFD-700 / Charles Anello, Ph.D.
HFD-700 / Satya D. Dubey, Ph.D.

C:\AAA\NDA\Alimta_NSCLC (21-677)\REVIEW\Report\Alimta.doc