Official Title: A Phase III, Multicenter, Placebo-Controlled, Double-Blind, Randomized Clinical Trial to Evaluate the Efficacy of Bevacizumab in Combination With Tarceva® (Erlotinib) Compared With Tarceva Alone for Treatment of Advanced Non−Small Cell Lung Cancer (NSCLC) After Failure of Standard First-Line Chemotherapy

NCT Number: NCT00130728

Document Date: SAP Version A1: 6-December-2006

STATISTICAL ANALYSIS PLAN

TITLE:	**A PHASE III, MULTICENTER, PLACEBO-CONTROLLED, DOUBLE-BLIND, RANDOMIZED CLINICAL TRIAL TO EVALUATE THE EFFICACY OF BEVACIZUMAB IN COMBINATION WITH TARCEVA® (ERLOTINIB) COMPARED WITH TARCEVA ALONE FOR TREATMENT OF ADVANCED NON–SMALL CELL LUNG CANCER (NSCLC) AFTER FAILURE OF STANDARD FIRST-LINE CHEMOTHERAPY**
PROTOCOL NUMBER:	OSI3364g
STUDY DRUG:	Bevacizumab Tarceva® (Erlotinib)
IND NUMBER:	BB-IND 7023
SPONSOR:	Genentech, Inc. 1 DNA Way South San Francisco, CA 94080-4990 U.S.A.
PLAN PREPARED BY:	[redacted], Ph.D. [redacted], M.S.
RECORDS RETENTION:	Genentech Central Records
DATE FINAL:	11 July 2005
DATE AMENDED:	6 December 2006

AMENDMENT SUMMARY TO THE OSI3364g
STATISTICAL ANALYSIS PLAN
6 December 2006 AMENDMENT

RATIONALE

The OSI3364g Statistical Analysis Plan has been amended to reflect changes to the study design made in the protocol amendments. The major changes are as follows:

- An interim efficacy analysis has been incorporated into the study design. The details of the interim efficacy analysis are now included.
- The study has been amended to include study sites from other countries. Region of study sites (United States vs. the rest of the world) is added as one of the factors to be assessed for potential differences in treatment effect in the exploratory efficacy analysis.
- The inclusion criteria have been expanded to allow patients with peripheral squamous cell carcinoma, patients with treated brain metastasis prior to enrollment, and patients requiring therapeutic anticoagulation with low molecular weight heparins or fondaparinux to be enrolled. Therefore, additional summaries of adverse events are added for these subgroups of patients.
- Potential sensitivity analyses are planned for overall survival if substantial treatment crossover to bevacizumab-class agents is observed.

Additional editorial changes have been made to improve clarity and consistency. All significant new information appears in italics. This amendment represents cumulative changes to the original Statistical Analysis Plan.

TABLE OF CONTENTS

Page

1. PROTOCOL SYNOPSIS

The protocol synopsis for Study OSI3364g (*Amendment 4, 30 November 2006*) is in Appendix A. For additional details, see the study flowcharts in Appendix B.

2. RANDOMIZATION ISSUES

Patients will be randomized to receive either Tarceva + placebo or Tarceva + bevacizumab using an interactive voice response system (IVRS) after written informed consent has been obtained and eligibility has been established. Patients may be randomized up to 5 days prior to study treatment initiation. A hierarchical dynamic randomization scheme will be used to ensure approximately equal sample sizes for the two treatment arms overall, within each of the four categories defined by baseline ECOG performance status (0/1 vs. 2) and smoking history (never vs. current/previous), within each sex (male vs. female), and within each study site. The detailed randomization scheme is provided in Appendix C.

The IVRS vendor, ClinPhone Inc. (Nottingham, UK), will develop the program to implement the hierarchical dynamic randomization. A Genentech study statistician and the independent Data Coordinating Center (DCC), Statistics Collaborative, Inc. (Washington, DC), will review the simulated randomization log and results prior to the activation of the IVRS. In addition, the DCC will check the treatment assignment every month during the study to ensure randomization is carried out correctly.

3. STATISTICAL METHODS

This study is a randomized, placebo-controlled, double-blind, parallel-group trial designed to evaluate the efficacy and safety of bevacizumab in combination with Tarceva relative to Tarceva monotherapy in second-line patients with advanced non–small cell lung cancer (NSCLC). *An interim efficacy analysis will be performed after approximately 280 deaths (67% of the total 417 deaths required for the final analysis) have been observed. If the interim analysis result is positive (i.e., crosses the O'Brien-Fleming boundary; see Section 3.7), the study will be unblinded. Otherwise, this study will be unblinded for analysis when the total number of required events (417 deaths; see Section 4) has been reached.* A data cutoff date will be determined when this number of events has occurred. The clinical database will

be thoroughly cleaned and frozen, the treatment assignment will be unblinded, analyses will be performed, and a clinical study report will be prepared.

Descriptive summaries of continuous data will present the group mean, standard deviation (SD), median, minimum, maximum, and sample size. Descriptive summaries of discrete data will present the number of patients as a frequency and as a percentage.

3.1 ANALYSIS OF STUDY CONDUCT

The number of patients who are randomized will be tabulated by center and by treatment arm. *Eligibility exceptions and p*rotocol deviations will be summarized by treatment arm. Patient disposition will be tabulated *by treatment arm*, and reasons for premature *study* discontinuation will be summarized *by treatment arm*.

Tarceva and bevacizumab exposure and treatment compliance will be summarized by treatment arm. Treatment compliance will be measured by dose intensity, which is defined as the total cumulative dose received divided by the total expected cumulative dose from the first to the last dose.

3.2 ANALYSIS OF TREATMENT GROUP COMPARABILITY

Treatment arms will be assessed for comparability with respect to demographic and baseline characteristics, including age; sex; race/ethnicity; weight; smoking status; histology; disease measurability; baseline ECOG performance status; number of disease sites; *history of treated brain metastases*; epidermal growth factor receptor (EGFR) expression status by immunohistochemistry (IHC); *EGFR gene copy number by fluorescent in situ hybridization (FISH)*; and EGFR and K-ras mutation status. No formal tests of hypotheses will be performed. The baseline value of any variable will be defined as the last available value prior to the first administration of study treatment.

3.3 EFFICACY ANALYSES

Unless otherwise specified, efficacy analyses will include all patients who are randomized and will be based on the treatment arm to which the patients have been randomized. For objective response and clinical benefit, only patients with measurable disease at baseline will be included in the analysis. For duration of response, only responders will be included in the analysis.

To manage the overall type I error rate at the 5% level, the fixed sequence testing procedure (Westfall and Krishen 2001) will be used to adjust for multiple statistical testing of the primary and key secondary efficacy endpoints. These endpoints will be tested in the following order:

- Overall survival
- Progression-free survival (PFS)
- Objective response
- Duration of objective response

The treatment difference in the primary endpoint (overall survival) will be tested at the two-sided 5% overall significance level using the O'Brian Fleming boundary (see Section 3.7). If this test result is positive at either the interim or the final analysis, the treatment difference in PFS will be tested at the 5% significance level. Only if the treatment effect on PFS is positive, the difference in objective response between the two treatment groups will be tested at the 5% level, and so on. If any test result is not positive, testing of the subsequent endpoints will not occur.

No adjustment to the α level will be made for the other analyses.

3.3.1 Primary Efficacy Endpoint

The primary efficacy endpoint is overall survival, defined as the time from the date of randomization (as entered into the IVRS) until the date of patient death from any cause. All deaths will be included, regardless of whether they occur during treatment or following treatment discontinuation. For patients who have not died, survival data will be censored at the date of last contact.

The two-sided log-rank test, stratified by the randomization stratification factors *except study site*, will be used to perform hypothesis testing for assessing the primary study objective at the 5% significance level. The randomization stratification factors are ECOG performance status (0/1 vs. 2), smoking history (never vs. current/previous), sex (male vs. female), and study site; however, because of the large number of study sites in this trial, study site will not be included in any efficacy analyses adjusted for randomization stratification factors. The stratification values reported on the Case Report Form (CRF) will be used in the analysis. If $>5\%$ of patients have discrepancies in stratification factors between the CRF and the IVRS, a sensitivity analysis stratified by values reported on the IVRS will be performed. Both analyses will be based on the treatment arm to which the patients have been randomized.

The Kaplan-Meier method will be used to estimate the median overall survival for each treatment arm. Cox proportional hazard models, using two models (with and without stratification by the randomization stratification factors), will be used to estimate the hazard ratio (i.e., the magnitude of treatment effect and 95% confidence interval [CI]).

3.3.2 Secondary Efficacy Endpoints

a. Progression-Free Survival (PFS)

PFS is defined as the time from randomization to documented disease progression or death on study *treatment*, whichever occurs earlier. Disease progression will be assessed by the investigator according to the Response Evaluation Criteria in Solid Tumors (RECIST). Patients with a global deterioration of health status requiring discontinuation of treatment without objective evidence of disease progression will also be considered to have disease progression (i.e., clinical progression will also be considered as an event "per RECIST"). Death on study *treatment* is defined as death from any cause within 30 days of the last dose of study treatment.

Data for patients without disease progression and for patients who have not died on study *treatment* will be censored at the time of the last tumor assessment (or, if no tumor assessments were performed after the baseline visit, at the time of randomization). Analysis methods are the same as those described for overall survival.

b. Objective Response

Objective response is defined as a partial or complete response determined on two consecutive occasions ≥4 weeks apart. Only patients with measurable disease at baseline will be included in the analysis of objective response. Patients without a post-baseline tumor assessment will be considered non-responders.

An estimate of the objective response rate and its 95% CI will be calculated for each treatment arm. The 95% CI for the treatment difference will also be calculated. The Mantel-Haenszel χ^2 test stratified by randomization stratification factors (ECOG performance status [0/1 vs. 2], smoking history [never vs. current/previous], and sex [male vs. female]) will be used to compare the treatment difference. An unadjusted Pearson χ^2 test result will also be provided as a sensitivity analysis.

c. Duration of Objective Response

Only patients with an objective response will be included in the analysis of duration of objective response. Duration of objective response is defined as the period from the date of the initial partial or complete response until the date of disease progression or death on study *treatment* from any cause. Methods for handling censoring and analysis are the same as described for overall survival and PFS.

d. Evaluation of the Relationship between Exploratory Markers and Efficacy Outcomes

To evaluate the effect of exploratory markers on efficacy outcome, efficacy outcomes will be summarized by treatment arm within each subgroup determined by exploratory markers. *Patients with missing or unevaluable marker status will be excluded from all analyses described in this section.* Efficacy outcomes considered for this analysis will include overall survival, PFS, and objective response *rate*. The Cox proportional hazard model will be used to evaluate the difference in treatment effect *on overall survival and PFS* across subgroups defined by a biomarker by testing the interaction effect of treatment and the biomarker. *The biomarker will be treated as the stratification factor in the Cox model to allow for different underlying baseline hazard functions between the two subgroups defined by the biomarker when testing the treatment by biomarker interaction. The logistic*

regression model will be used to evaluate the difference in treatment effect on objective response rate across subgroups defined by a biomarker by testing the interaction effect of treatment and the biomarker. In addition, *the Cox model or logistic regression model* will be used to assess the clinical benefit in each subset defined by the biomarkers. The analyses *described in this section* will not *be stratified by randomization stratification factors unless substantial imbalances are observed* because of *an anticipated* small sample size in some subsets.

Markers to be considered include EGFR expression by IHC using the Dako EGFR PharmDx™ immunohistochemistry kit, EGFR gene copy number by fluorescence in situ hybridization (FISH) using the Vysis LSI® EGFR SpectrumOrange™/CEP® 7 SpectrumGreen™ kit, and the presence or absence of somatic gene mutations of the EGFR pathway by DNA sequencing. A positive EGFR expression by IHC will be defined as having $\geq 10\%$ of tumor cells staining for EGFR. A high copy number of the EGFR gene will be defined as high polysomy (greater than four gene copies in $>40\%$ of cells) or amplification (gene/chromosome >2, or >15 gene copies in $>10\%$ of cells). The EGFR gene will be considered wild type if no mutations are detected in exons 18–21.

Other markers relevant for the EGFR or VEGF pathway may also be explored.

3.3.3 Exploratory Efficacy Endpoints

a. Effect of Risk Factors

The effects of demographic and baseline characteristics on treatment comparisons for overall survival, PFS, and response rate will be examined as exploratory analyses. It is expected that many of the subgroups defined by these factors will be small and will not have adequate power to detect a treatment difference.

The risk factors to be considered include the following:

- Sex: male, female
- Baseline ECOG performance score: 0–1, 2
- *Sex by baseline ECOG performance score strata*
- Smoking history: never smoked, smokers
- *Age group: <65 years, ≥65 years*

- *Race/ethnicity*
- Best response to prior therapy: CR/PR (complete response/partial response), SD (stable disease), PD (progressive disease)
- Histology: adenocarcinoma, adenocarcinoma with BAC (bronchioloalveola carcinoma) or BAC-like features
- Time from initial NSCLC diagnosis to randomization: <6, 6–12, >12 months
- *Region of study sites: sites in the United States, sites in the rest of the world*

For overall survival and PFS, the Kaplan-Meier estimated median time will be summarized by treatment arm for each of the subgroups defined above along with a hazard ratio (treatment:control) estimated by unstratified Cox regression displayed on a Forest plot (Lewis and Clarke 2001). The difference in treatment effect across subgroups will be assessed using Cox regression to test the interaction effect between treatment and each of the above factors separately. *The risk factor will be treated as the stratification factor in the Cox model to allow for different underlying baseline hazard functions among the subgroups defined by the risk factor when testing the treatment by factor interaction*. In addition, the overall treatment difference with adjustment for each factor will be evaluated using Cox regression with *risk* factor and the randomization stratification factors *as covariates in the model*.

For objective response, the response rate will be summarized by treatment arm for each of the subgroups defined above along with the odds ratio (treatment:control) estimated by logistic regression displayed on a Forest plot. The difference in treatment effect across subgroups will be assessed using logistic regression by testing the interaction effect between treatment and each of the above factors separately. In addition, the overall treatment difference with adjustment for each factor will be evaluated using logistic regression with adjustment for the factor and the randomization stratification factors.

b. Clinical Benefit

Clinical benefit will be defined as objective response or stable disease maintained for at least *6* weeks from the start of treatment. Only patients with measurable disease at baseline will be included in the analysis of clinical benefit. The clinical benefit rate will be analyzed as described for the objective response rate.

c. Sensitivity Analysis

Patients can receive any subsequent treatments during survival follow-up after the end of study treatment to warrant best supportive care. Consequently, the potential survival benefit of the treatment relative to control may be obscured by the potential benefit patients receive from subsequent treatments. For example, treatment effects may be masked if a substantial proportion of patients receive bevacizumab or agents in the same class as one of the subsequent treatments (i.e., crossover to bevacizumab-class). To address this issue, sensitivity analyses of overall survival will be performed if substantial crossover is observed.

The analyses may include but not limited to the following:

1. *Patients receiving bevacizumab-class agents in the subsequent treatments will be censored at the initial time of receiving the first subsequent bevacizumab-class agent. Analysis methods are otherwise the same as those described for overall survival in Section 3.3.1.*
2. *Overall survival analysis will be performed in patients who did not receive any bevacizumab-class agents in subsequent treatments. Analysis methods are the same as those described for overall survival in Section 3.3.1.*

3.4 PHARMACOKINETIC ANALYSES

The purpose of this pharmacokinetic study is to characterize the pharmacokinetic behavior of erlotinib and bevacizumab in a subset of patients with previously treated advanced NSCLC.

Intensive pharmacokinetic sampling will be conducted at selected study sites to evaluate the drug–drug interaction in pharmacokinetics between erlotinib and bevacizumab. To evaluate the effect of bevacizumab on erlotinib pharmacokinetics in the two treatment arms (Tarceva vs. Tarceva+bevacizumab), a subset of 17 evaluable patients will be needed. A patient will be considered evaluable for erlotinib pharmacokinetic analysis if he or she has a full profile of erlotinib (i.e., he or she has sufficient samples to determine the pharmacokinetic outcome measures). In addition, to evaluate the effect of erlotinib on bevacizumab pharmacokinetics, peak and trough concentrations will be obtained from those patients participating in the intensive pharmacokinetic study, and the concentration levels will be compared with historical concentration data. Subjects who received bevacizumab and had at least one peak or one trough

concentration at any one of the sampling timepoints (Days 21, 42, 63, and 84) will be included in this analysis.

Pharmacokinetic analyses will be performed separately for bevacizumab and erlotinib.

3.4.1 Assessment of the Effect of Bevacizumab on Erlotinib Disposition

Plasma sampling for erlotinib in both arms will be performed at Week 3, when erlotinib is at steady state, and following the second dose of bevacizumab or placebo. Samples will be drawn at the following timepoints: Day 21, pre−Tarceva dose and 2, 4, 6, 8, 10, and 12 hours post−Tarceva dose; Day 22, pre−Tarceva dose. In addition, because α-1-acid glycoprotein (AAG) is a significant covariate in erlotinib clearance, samples for AAG levels will be drawn on Day 21 pre−Tarceva dose. Actual doses and sample collection times will be used in the pharmacokinetic analysis.

The plasma concentrations of erlotinib and its O-demethylated metabolite (OSI-420) versus actual time for each subject will be presented graphically in the raw and $\log_{10}$-transformed scale, respectively. OSI-420 is the primary and active metabolite of erlotinib produced by metabolism by the cytochrome P-450 system. The mean (+standard error of mean [SEM]) plasma erlotinib and OSI-420 concentrations versus scheduled time will also be presented graphically by treatment. Note that concentrations below the limit of quantification will be imputed using one-half the limit of quantification in all summaries and analyses.

Erlotinib and OSI-420 maximum concentration (C_{max}), time at which the maximum concentration was achieved (t_{max}), maximum concentration at steady state ($C_{max, ss}$), area under the erlotinib time−concentration curve ($AUC_{0-\tau}$), minimum concentration at steady state ($C_{min, ss}$), and apparent clearance (CL/F) (erlotinib only) for each treatment group will be estimated by the non-compartmental approach using a validated software program (WinNonlin®, Version *5.5.1*, Pharsight Corp., Mountain View, CA). The AUCs will be calculated using the linear trapezoidal rule up to the time at which the maximum concentration was achieved (t_{max}) and using the log trapezoidal rule after t_{max}. $AUC_{0-\tau}$ will be calculated from the time of dosing (0 hour) to the last measurable concentration (C_τ).

Descriptive statistics, including mean, SD, median, range, and geometric mean for each pharmacokinetic parameter, will be tabulated by treatment group.

Only patients who have estimates of $AUC_{0-\tau}$ and C_{max} will be included in the assessment of drug interaction of bevacizumab with erlotinib. The drug–drug interaction will be evaluated by calculating the 90% CIs of the ratios of $AUC_{0-\tau}$ and C_{max} between the two treatment arms. The analysis of variance (ANOVA) approach will be used to obtain the least squares (LS) means and standard errors (SE) of $AUC_{0-\tau}$ and C_{max} in natural log scale. The ratio of $AUC_{0-\tau}$ and C_{max} between the two treatment groups will be estimated by taking the anti-log of the difference in the LS mean between the two arms. The 90% CI will be obtained by taking the anti-log of the 90% CI for the mean difference in log scale. That is, the 90% CI is the anti-log of

$$\text{LS mean}_T - \text{LS mean}_{TA} \pm t_{.95}[df] \bullet SE(\text{LS mean}_T - \text{LS mean}_{TA}),$$

where

LS mean_T and LS mean_{TA} are the LS mean for the Tarceva alone and Tarceva + bevacizumab arms (log scale), respectively,

$t_{.95}[df]$ is the 95th percentile of the Student t-distribution with degrees of freedom for overall error from the ANOVA model, and

$SE(\text{LS mean}_T - \text{LS mean}_{TA})$ is the SE of the contrast term for the difference in LS means.

3.4.2 Assessment of the Effect of Erlotinib on Bevacizumab Disposition

Because there is no arm in which bevacizumab is administered alone (in the absence of erlotinib), the assessment of the effect of erlotinib on bevacizumab pharmacokinetics will be different from the approach described above. The effect of erlotinib on bevacizumab will be assessed for all patients who have received bevacizumab and have had at least one peak or one trough concentration at any one of the sampling timepoints (Days 21, 42, 63, and 84).

Bevacizumab concentrations in the Tarceva + bevacizumab arm will be compared with historical concentration data used in a population pharmacokinetic analysis in cancer patients receiving bevacizumab in Phase I–III studies (Report 03-0324-1751 [Population Pharmacokinetics of Bevacizumab: Structural Model Identification, Mean Population Pharmacokinetic Parameter

Estimation, and Covariate Analysis], submitted as part of the bevacizumab Biologics License Application on 25 September 2003). Specifically, the effect of erlotinib on bevacizumab pharmacokinetic disposition will be evaluated as follows:

- C_{max} and C_{min} at Days 21, 42, 63, and 84 in patients from this study will be compared with historically observed values in patients from the population pharmacokinetic database on the same dose and schedule of bevacizumab.
- The mean and SD of C_{max} and C_{min} at Days 21, 42, 63, and 84 in patients from this study will be $_{calculated}$ and compared with historical data from patients from the population pharmacokinetic database who were given 15 mg/kg bevacizumab every 3 weeks. The individual C_{max} and C_{min} versus actual time for each patient and the mean (SEM) concentrations versus scheduled time for the data from this study and the historical data will be presented graphically.
- C_{max} and C_{min} at Days 21, 42, 63, and 84 will be compared with simulated concentrations at those timepoints based on the population pharmacokinetic model from Report 03-0324-1751.
- First, 500 patients will be drawn with replacement from the bevacizumab population pharmacokinetic database. The simulated C_{max} and C_{min} at Days 21, 42, 63, and 84 for these 500 patients following administration of 15 mg/kg bevacizumab every 3 weeks will be obtained based on the population pharmacokinetic model from Report 03-0324-1751, with the random and residual effects simulated from the normal distribution. Then, the mean and SD of the predicted C_{max} and C_{min} at each timepoint for these 500 patients will be calculated and informally compared with the observed concentration data from this study. No formal statistical comparisons will be made.

3.5. SAFETY ANALYSES

Adverse events, laboratory test results, and changes in blood pressure will be summarized by treatment arm. Patients who receive any amount of study treatment will be included in the safety analyses. Safety results will be summarized based on the treatment that patients actually receive.

3.5.1 Adverse Events

Verbatim descriptions of treatment-emergent adverse events will be mapped to Medical Dictionary for Regulatory Activities (MedDRA) thesaurus terms and graded according to the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE), Version 3.0.

The incidence in patients of the following events will be summarized for each treatment group by system organ class and preferred term. For each patient's adverse events, the maximum severity recorded will be used in the summary.

- All adverse events
- Grade 3–4 and Grade 5 adverse events
- All deaths
- Serious adverse events
- Adverse events leading to discontinuation of study drug
- Adverse events by subgroup: incidence of all adverse events summarized by age at baseline (<65, ≥ 65 years old), sex (male, female), and race/ethnicity
- Adverse events in a special population: the incidence of all adverse events in patients with squamous cell carcinoma, patients with treated brain metastases prior to enrollment, and patients requiring low molecular weight anticoagulant

3.5.2 Laboratory Data

Clinical laboratory tests will be performed at local laboratories. Laboratory toxicities will be defined based on local laboratory normal ranges and the NCI CTCAE, Version 3.0. For each laboratory parameter, the toxicity grade at baseline and the worst toxicity grade during the treatment phase will be summarized by treatment group. The worst toxicity grade shift from baseline for white blood count, absolute neutrophil count, platelet count, and SGOT and SGPT levels will be summarized by treatment arm.

3.5.3 Vital Signs

Blood pressure will be measured at baseline and every 3 weeks during the study treatment phase. Changes in blood pressure over time will be summarized graphically and in tabular format by treatment arm.

3.6 MISSING DATA

For overall survival, patients who are lost to follow-up will be treated as censored on the date of last contact.

For PFS, patients who do not have disease progression and who have not died on study *treatment* will be treated as censored at the time of the last tumor assessment (or, if no tumor assessments were performed after the baseline visit, at the time of randomization).

For objective response and clinical benefit, patients without a post-baseline tumor assessment will be considered non-responders.

For duration of objective response, patients who do not have disease progression and who have not died on study *treatment* will be treated as censored at the time of the last tumor assessment.

3.7 INTERIM ANALYSES

A Data Monitoring Committee (DMC) will monitor safety and will meet periodically to review safety summaries prepared by the external statistical DCC.
The detailed interim safety analysis plan and the role and responsibilities of the DMC members are described in the separate Charter for the Data Monitoring Committee. Members of the DMC will be external to Genentech and will follow the charter. Study personnel will remain blinded to study results until final study unblinding.

An interim efficacy analysis is planned and will be performed by the DCC after approximately 280 deaths (67% of the total required deaths for final analysis) have been observed. The overall survival will be tested at the significance level determined using the Lan-DeMets α-spending function with an O'Brien-Fleming boundary of 0.0124 at 67% event time and 0.0462 at the final analysis, so that the overall type I error rate will be maintained at the 0.05 level. The two-sided log-rank test, stratified by the randomization stratification factors except study site, will be used to perform hypothesis testing for assessing the primary study objective. If the overall survival result is positive, the study will be unblinded and the primary analysis will be performed. Otherwise, the study will continue until 417 deaths have occurred.

4. DETERMINATION OF SAMPLE SIZE

This is a Phase III study to evaluate the efficacy and safety of bevacizumab in combination with Tarceva relative to Tarceva alone in advanced NSCLC. A 33% improvement in median survival is considered a clinically significant outcome. To calculate the number of deaths required for this study, the following assumptions were made:

- Two-sided log-rank test
- *83*% power at the 5% significance level
- Hazard ratio of bevacizumab + Tarceva versus control (Tarceva alone) of 0.75 corresponding to a 33% improvement in median overall survival from 8 to 10.67 months
- *An interim efficacy analysis will be performed when 67% of the required deaths for final analysis have occurred. The significance level will be determined using the Lan-DeMets α-spending function with an O'Brien-Fleming boundary.*
- *Five percent of patients will withdraw consent for survival follow-up or will be lost to follow-up.*

Median overall survival for the control arm is hypothesized based on data from Study BR.21, taking into account the difference in patient population.

With these assumptions, *417* deaths are required. Assuming an accrual rate of *20* patients per month, 650 patients enrolled over *35* months *(accounting for staggered accrual)* and followed for an additional *2* months will provide *417* deaths.

For the subpopulation undergoing intensive pharmacokinetic analysis, 17 evaluable patients will be needed in each treatment arm. A patient will be considered evaluable for pharmacokinetic analysis if he or she has a full profile of erlotinib samples and at least two bevacizumab samples. This sample size is estimated to obtain 81% power for evaluating the equivalence of AUC between the two treatment arms with an equivalence limit of 0.67–1.5, assuming that the true mean log (natural log) of the AUCs of the two treatment arms is identical, that the AUC interpatient coefficient of variation is ~40% (based on the estimates from Report 034-0143-1219: Population Pharmacokinetics of Tarceva), and that data will be analyzed in the natural log scale using two one-sided t-tests for differences in mean AUC at the 5% significance level for each test. Additional patients may need to be added to obtain 34 evaluable patients.

5. TABLE OF CONTENTS OF PLANNED TABLES, LISTINGS, AND GRAPHS

Study Population and Disposition Data

Table 14.1/1: *Enrollment by Investigator (Randomized Patients)*

Table 14.1/2: *Patient Status and Reasons for Study Discontinuation (Randomized Patients)*

Table 14.1/3: *Reasons for Discontinuation of Treatment Phase (Randomized Patients)*

Table 14.1/4: *Reasons for Early Discontinuation of Bevacizumab/Placebo (Treated Patients)*

Table 14.1/5.1 *Patient Eligibility: Description of Inclusion and Exclusion Criteria*

Table 14.1/5.2 *Patient Eligibility: Inclusion and Exclusion Criteria Not Met (Randomized Patients)*

Table 14.1/6 *Number and Percentage of Patients Using Concomitant Medications on Study Treatment (Treated Patients)*

Efficacy and Pharmacokinetic Data

Table 14.2/1: *Demographics and Baseline Characteristics (Randomized Patients)*

Table 14.2/2: *Patient Compliance with Study Treatment (Treated Patients)*

Table 14.2/3: *Duration of Study Treatment (Treated Patients)*

Table 14.2/4: *Overall Survival (Randomized Patients)*

Table 14.2/5: *Progression-Free Survival (Randomized Patients)*

Table 14.2/6: *Objective Response (Patients with Measurable Disease at Baseline)*

Table 14.2/7: *Duration of Objective Response (Patients with an Objective Response)*

Table 14.2/8: *Overall Survival by Biomarker (Randomized Patients)*

Table 14.2/9: *Progression-Free Survival by Biomarker (Randomized Patients)*

Table 14.2/10: *Objective Response by Biomarker (Patients with Measurable Disease at Baseline)*

Table 14.2/11: *Overall Survival by Risk-Factor Subgroup (Randomized Patients)*

Table 14.2/12: *Progression-Free Survival by Risk-Factor Subgroup (Randomized Patients)*

Table 14.2/13: *Objective Response by Risk-Factor Subgroup (Patients with Measurable Disease at Baseline)*

Table 14.2/14: *Overall Survival Adjusted for Risk Factors (Randomized Patients)*

Safety Data

6. REFERENCES

Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. BMJ 2001;322:1479–80.

Westfall PH, Krishen A. Optimally weighted, fixed sequence and gatekeeper multiple testing procedures. J Stat Planning Inference 2001;99:25–40.