# A Double-blind, Placebo-controlled, Phase 3 trial of Exemestane or Fulvestrant With or Without Seribantumab in Postmenopausal Patients with Hormone Receptor-positive, Heregulin Positive (HRG+), HER2 negative Metastatic Breast Cancer Whose Disease Progressed After Prior Systemic Therapy

**Sponsor:**

**Merrimack Pharmaceuticals, Inc.**
One Kendall Square
Suite B7201
Cambridge, MA
02139

# Signature Page

Upon review of this document, the undersigned approves this version of the Statistical Analysis Plan, which authorizes that the content is acceptable based on the current planned protocol.

| | Name | Signature | Date |
|---|---|---|---|
| **Author:** | Ty McClure | | |
| **Position** | Senior Director Biostatistics | | |
| **Company** | Merrimack Pharmaceuticals | | |
| | | | |
| **Approved by:** | Akos Czibere | | |
| **Position** | Senior Medical Director | | |
| **Company** | Merrimack Pharmaceuticals | | |

MERRIMACK PHARMACEUTICALS

**TABLE OF CONTENTS**

# I. Contents

MERRIMACK PHARMACEUTICALS

# ABBREVIATIONS

| | |
|---|---|
| AE | Adverse event |
| AUC | Area under the curve |
| BSA | Body Surface Area |
| CBR | Clinical benefit rate |
| CI | Confidence Interval |
| Cmax | Maximum serum concentration |
| CR | Complete Response |
| CRF | Case report form |
| CSR | Clinical study report |
| CT | Computed tomography |
| CTCAE | Common terminology criteria for adverse events |
| DMC | Data Monitoring Committee |
| DOR | Duration of Response |
| DSMB | Data Safety Monitoring Board |
| ECG | Electrocardiogram |
| ECHO | Echocardiogram |
| eCRF | Electronic case report forms |
| EDC | Electronic data capture |
| FDA | Food and Drug Administration |
| HRG | Heregulin |
| ICH | International conference on harmonization |
| ITT | Intent-to-Treat |
| IWRS | Interactive web response system |
| LVEF | Left ventricular ejection fraction |
| MedDRA | Medical Dictionary for Regulatory Activities |
| Mth | Month |
| MUGA | Multiple gated accession scan |
| NCI | National Cancer Institute |
| NE | Not Evaluable |
| ORR | Objective response rate |
| OS | Overall Survival |
| PD | Progressive disease |
| PFS | Progression free survival |
| PK | Pharmacokinetic |
| PR | Partial response |
| PT | Preferred term |
| QTcF | Fridericia's QT correction |
| RECIST | Response Evaluation Criteria In Solid Tumors |
| SAE | Serious adverse event |

| | |
|---|---|
| SAP | Statistical Analysis Plan |
| SD | Stable Disease |
| SOC | System Organ Class |
| SRB | Safety Review Board |
| Tmax | Time to maximum serum concentration |
| t 1/2 | Terminal elimination half-life |
| TFL | Tables, Figures, Listings |
| TTF | Time to treatment failure |
| ULN | Upper limit of normal |
| Vdss | Volume of distribution at steady state |
| Wk | Week |

# 1. INTRODUCTION

This is the Statistical Analysis Plan (SAP) for protocol seribantumab-02-03-10.

## 1.1 Protocol Changes and Clarifications

No previous signed off SAP versions have been written as of the current sign-off. The following are clarifications for protocol analysis.

## 1.2 Study Design

This study is a double-blind, placebo-controlled, phase 3 trial of exemestane or fulvestrant with or without seribantumab in postmenopausal patients with hormone receptor-positive, heregulin positive (HRG+), HER2 negative metastatic breast cancer whose disease progressed after prior systemic therapy.

## 1.3 Study Objectives

The primary objective of this study is to determine whether the combination of seribantumab plus exemestane or seribantumab plus fulvestrant is more effective than placebo + exemestane or placebo + fulvestrant based on investigator assessed Progression Free Survival (PFS) in HRG positive patients (defined as HRG ISH score of $\geq$ 1+).

The secondary objectives of this study are as follows:

- To determine whether the combination of seribantumab plus exemestane or seribantumab plus fulvestrant is more effective than placebo + exemestane or placebo + fulvestrant in HRG positive patients for the following clinical outcome parameters:
  - Time to Progression (TTP)
  - Objective Response Rate (ORR) based on RECISTv1.1
  - Overall Survival (OS)
- To compare the time from randomization to subsequent chemotherapy between the seribantumab containing arms and the comparator arms of the study To describe the safety profile of seribantumab in combination with exemestane or fulvestrant
- To describe the safety profile of seribantumab in combination with exemestane or fulvestrant
- To assess health-related quality of life (HRQOL) using EuroQol 5-Dimension (EQ-5D-5L), Functional Assessment of Cancer Therapy – Breast (FACT-B) and Brief Pain Index (BPI) questionnaires
- To characterize the pharmacokinetic (PK) profile of seribantumab when given in combination with exemestane or fulvestrant and of exemestane or fulvestrant when given in combination with seribantumab
- To assess the effect of seribantumab on ECG parameters

The exploratory objectives of this study are as follows:

- To assess the correlation for HRG expression between fresh tissue biopsies and archival samples where available
- To estimate resources required for hospitalization

# 2. Analysis Sets

## 2.1 Intent-to-Treat (ITT)

The ITT population will include all patients who are randomized, with study drug assignment via an Interactive Web Response System (IWRS), irrespective of whether patients receive the intended study drug or a different drug than

allocated through randomization. ITT patients with central lab HRG positive (1+, 2+, 3+, 4+) results will be considered the primary efficacy population. Patients with missing or indeterminate HRG or HRG negative will not be part of the ITT primary efficacy population.

## 2.2 Safety Population (SAF)

This will include all patients in the study receiving at least 1 dose of study drug. These patients will be defined as the safety population.

## 2.3 Pharmacokinetic (PK)

This population includes all SAF patients with at least 1 on-treatment PK assessment.

# 3. ENDPOINTS

## 3.1 Progression-Free Survival – Investigator Assessed

Investigator Assessed (IA) tumors assessments are those determined by the investigator. PFS is defined as the number of months from the date of randomization to the date of death or documented (objective) disease progression via RECISTv1.1, whichever occurs earliest. Length of PFS will be calculated in days and reported in months.

[Progression/Death/Censor Date - Randomization Date + 1] / 30.4375.

The tumor assessment (i.e., scan dates) will be used for progression/censor date not the date corresponding to the determination of overall response. When multiple dates occur for assessing target, non-target, and new lesions the following rule will be applied:

- If documented disease progressions, the ***first*** available date will define the tumor assessment date
- If censored, the ***last*** available date will define the tumor assessment date.

Per protocol tumor assessments are scheduled to be performed every 8 weeks (± 7 days) from day of first dose until investigator-assessed progression, in addition to a 30-day follow-up.

**Censorship**: In general, patients are censored at the date of the last tumor assessment that verified lack of disease progression if they are last known to be alive, on-treatment or within 30 days following treatment discontinuation, and progression-free. Censoring is explicitly described below and in table 1. Patients that do not experience progression or death by the time of analysis (data cut-off) will be censored at the date of last tumor assessment defined by any overall response other than "not evaluable".

- Patients with an inadequate/incomplete baseline tumor assessment will be censored at the date of randomization
- Patients with no tumor assessment after dosing will be censored at the date of randomization unless death occurred prior to the first planned assessment (in which case the death is an event)
- Patients with at least one on-study disease assessment who discontinue treatment without documented disease progression or death are censored at the date of the last tumor assessment

  There are two exceptions.

  - Documented progression or death ≤ 30 days after treatment discontinuation then progression or death is an event
  - A new anti-cancer treatment is started prior to documented progression and ≤ 30 days after treatment discontinuation, then censorship is at the date of the last tumor assessment that verified lack of disease progression prior to the new treatment
- Patients with documentation of progression or death after an unacceptably long interval (i.e., 2 or more consecutively missed or indeterminate/not evaluable assessments) since the last tumor assessment will be censored at the date of last tumor assessment documenting no progression

**Table 3.1 Documented Disease Progression and Censoring**

| Situation | Date of | Outcome |
|---|---|---|
| Inadequate baseline tumor assessment | Randomization | Censored |
| No on-treatment assessment and no death prior to first scheduled assessment | | |
| Death or documented disease progression after ≥ 2 consecutively missed and/or not evaluable scheduled tumor assessments | Last tumor assessment documenting no disease progression | |
| Alive, on-treatment, and no documented disease progression | Last tumor assessment documenting no disease progression | |
| Documented disease progression (on, prior to, or ≤ 30 days after treatment discontinuation) | First tumor assessment documenting disease progression | Progression (Event) |
| New anticancer treatment ≤ 30 days after treatment discontinuation without disease progression | Last tumor assessment documenting no disease progression prior to starting the new anticancer treatment | Censored |
| Treatment discontinuation due to toxicity, undocumented disease progression, or any other reason | Last tumor assessment documenting no disease progression prior to discontinuation | Censored |
| Death prior to first scheduled assessment | Death | Death (Event) |
| Death without documented disease progression and ≤ 30 days after treatment discontinuation | Death | Death (Event) |

Note: Follow-up scans can be performed up to 4 weeks from last treatment dose date via protocol

## 3.2 Overall Survival

Overall Survival (OS) is defined as the time from randomization to the date of death for any cause. In the absence of confirmation of death, survival time will be censored at the last date the patient is known to be alive. Length of OS will be calculated in days and reported in months.

[Death/Censor Date - Randomization Date + 1] / 30.4375.

## 3.3 Progression-Free Survival – Independent Radiologic Review

Independent radiologic review (IRR) will be performed using images from all available tumor assessments. PFS is defined as the number of months from the date of randomization to the date of death or documented (objective) disease progression via RECISTv1.1, whichever occurs earliest. Length of PFS will be calculated in days and reported in months.

[Progression/Death/Censor Date – Randomization Date + 1] / 30.4375.

Censorship rules for PFS per IRR will be defined the same as the investigator assessed.

## 3.4 Time to Progression

Time to Progression (TTP) will be derived using the investigator-assessed tumor assessment data. Censorship will occur for patients without documented disease progression. The last tumor assessment documenting no PD will be

used as the date of event/censor. If death occurs prior to first tumor assessment the date of death will be used as the date of event/censor. Length of TTP will be calculated in days and reported in months.

[Progression/Censor Date – Randomization Date + 1] / 30.4375.

## 3.5 Objective Response Rate

Objective Response Rate (ORR) will be derived using the investigator-assessed tumor assessment data. Best Overall Response (BOR) is recorded from randomization until documented disease progression or death. ORR is defined as the proportion of patients with a BOR characterized as either a Complete Response (CR) or Partial Response (PR) as defined according to RECISTv1.1 guidelines relative to the total number of evaluable patients. Per RECISTv1.1 guidelines, patients with documented PR or CR will not need confirmatory tumor assessments to document response. Patients without baseline or post-baseline tumor assessments will be considered non-evaluable for this analysis. The following table describes the responder/non-responders for this analysis.

**Table 3.5 ORR Determination**

| Criteria | Responder/Non-responder |
|---|---|
| No Baseline tumor assessment | Not Evaluable |
| No Post-baseline tumor assessment AND were treatment failure prior to BOR determination due to disease progression, symptomatic deterioration, toxicity, death, or new anticancer treatment | Non-Responder |
| BOR: Partial Response, Complete Response<br>AND were not treatment failure prior to BOR determination due to disease progression, symptomatic deterioration, toxicity, death, or new anticancer treatment | Responder |
| BOR: Stable Disease, Progressive Disease, Not Evaluable/Indeterminate | Non-responder |

## 3.6 Safety Endpoints

Safety profile will be characterized by type, frequency, and severity of adverse events, coded using the MedDRA coding system, as graded by NCI Common Toxicity Criteria for Adverse Events version 4.0 (NCI CTCAEv.4), timing and relationship to treatment, laboratory abnormalities, vital signs, and left ventricular imaging data.

### 3.6.1 Treatment-Emergent Adverse Events (TEAE)

Adverse events will be considered treatment-emergent if:

- The event occurs for the first time after the start of study treatment and on or before 30 days after final dose of study treatment and was not observed prior to start of treatment
- The event was observed prior to the start of treatment but increased in NCI CTCAE v4 grade during study treatment

The eCRF will capture TEAEs on the adverse event page, while those events occurring pre-treatment will be captured on medical history.

### 3.6.2 Laboratory, ECG, and Vital Signs Data

Assessments will be assigned to cycles based on the collection date of the sample relative to the start dates of the cycles from the study drug administration page.

Baseline assessments will be considered as those data collected:

- Within 28 days prior to first day of study drug and
- The closest non-missing data value assessment prior to the first dose date

## 3.7 Stratification Factors

Randomization will be stratified based on the following prognostic factors:

- Prior chemotherapy for locally advanced or metastatic disease (yes, no)
- Chemotherapy backbone (exemestane or fulvestrant)
- Geographic Region (US, Non-US/Non-Asia, Asia)

## 3.8 Covariates

Baseline evaluations are those collected within 28 days prior to the start of treatment. If more than one baseline evaluation is available, select the closest non-missing value prior to first dose. The closest non-missing observation may be selected on the day of cycle 1 day 1 if study procedures are indicated prior to dosing.

The following baseline characteristics will be included as covariates which may be included in model testing exercises: Age, gender, race, ethnicity, primary diagnosis and time from primary cancer diagnosis to randomization, time since last prior systemic treatment, investigative site, and others may be identified.

## 3.9 Biomarkers

Baseline serum and tissue samples will be collected. Continuous and categorical type of biomarkers may be available. For continuous biomarkers, the value at and above the median will be used to determine the high group, while the low group is below. Categorical parameters will be classified into a maximum of three groups. Additional analyses will be considered exploratory in nature. Samples will be used to assess additional clinically relevant biomarkers, including but not limited to expression levels of the ErbB3 receptor, cMET, IGF-1R, and FGF receptor tyrosine kinases and will include assessment of their respective ligands (e.g. HGF, FGF-2, IGF-1, and IGF-2 mRNA or protein).

The only biomarker detailed in this SAP will be HRG positive as per the primary efficacy population analysis.

# 4. HANDLING OF MISSING DATA

## 4.1 General Conventions

Dates will be presented in listings as DDMONYYYY. Listings will be presented as the collected or raw data with missing information presented as such: - - MONYYYY for missing day, DD - - - YYYY for missing month, and DDMON - - - - for missing year. Missing day values will have the first day of that month imputed if calculating a start date, while stop dates will have the last day of that month. Missing month and/or year will require a review of patients with similar events. The month and/or year will be selected such that the mean value is approximated using that particular value(s).

## 4.2 Tumor Assessments

Inadequate tumor assessment does not allow for a RECIST response determination and will be designated as Not Evaluable target response evaluation. An inadequate assessment will be defined as those with at least 1 missing measurement or non-measurable target lesion or baseline tumor assessments outside the screening evaluation window (i.e., after first dose date and prior to -28 days from first dose date). If a target lesion is too small to measure and does not have unequivocal complete disappearance, a default of 5mm will be assigned.

Different imaging of lesions should not be performed. Those tumor assessments that do not vary based on different imaging will be allowed. However, imaging variations that result in PR or CR based on the inconsistency of imaging will be reviewed carefully and may result in confirmation of that result at the following visit.

If non-target disease is not assessed and a patient is considered for a CR target response the overall disease response will be PR.

## 4.3 Safety Data

Missing adverse event, vital sign, ECG, or laboratory data will be considered missing in summary analyses.

## 4.4 Pharmacokinetic Parameters

Missing values will be considered missing for summarizations and analyses. NC will be set to missing. In all data presentations other than listings, concentrations below the limit of quantification (BLQ) will be set to zero. The listings will report BLQ as < LLQ or lower limit of quantification.

Summary tables and plots will be calculated by setting missing concentrations if a concentration has been collected as ND (i.e., not done) or NS (i.e., no sample) or NC (i.e., not computable) or a deviation in sampling time is of sufficient concern or a concentration has been flagged as a data anomaly. Individual observations with known to bias the plasma concentration due to an unexpected event will be described and not included in summarizations or analysis.

## 4.5 Biomarkers

Missing or indeterminate HRG values considered as HRG positive will be not be included as part of the primary efficacy analysis. Patients with missing baseline biomarker assessments will not be used in analysis that will include baseline biomarker information. No missing data will be imputed.

# 5. STATISTICAL METHODS

## 5.1 Statistical Hypothesis and Determination of Sample Size

The primary efficacy analysis will be to compare PFS distributions of seribantumab in combination with physician's choice of fulvestrant and/or exemestane versus control (fulvestrant and/or exemestane alone). The study is designed to test superiority with the null hypothesis of no differences in progression-free survival distributions using investigator assessed disease progression between the treatment and control arm.

The primary PFS analysis will be performed using a stratified log-rank test, where stratification factors are prior chemotherapy use (yes, no), chemotherapy backbone (exemestane or fulvestrant), and geographic region (US, Non-US/Non-Asia, Asia). The stratified log rank test will be performed using a one-sided significance level at 0.025.

Interim analysis will be conducted when reporting approximately 50% (approximately 16 months from first patient) of the final PFS events have been observed in the primary efficacy population. Type II error will be preserved using a Lan-DeMets (non-binding) beta spending function approach with an O'Brien-Fleming boundary. Non-binding for the futility analysis implies that the futility boundary will be constructed in such a way that it can be overruled if desired by the DMC without inflating the type I error rate and without decreasing power. The interim will be a futility analysis only. The planned interim analysis is detailed in section 5.1.1.

Since the boundary is dependent on the number of PFS events the actual boundary used will be re-calculated based on the number of actual PFS events achieved at the time of the interim analysis. The DMC will be responsible for evaluating the interim analysis and making a recommendation about early termination due to observed study results as outlined in the DMC charter.

Approximately 195 PFS events are be required to have at least 95% power to detect a 43% risk reduction (mPFS: 4 v 7; HR $\leq$ 0.57) for patients receiving the combination of seribantumab plus exemestane or fulvestrant versus exemestane or fulvestrant alone, using a one-sided, stratified log-rank test at a significance level of 0.025, assuming a 2:1 ratio. An accrual of 24 months in length to obtain the required number of 195 PFS events is estimated to be achieved at approximately 30 months with a total sample size of 286 patients.

### 5.1.1 Safety Review and Interim Analysis

An independent external DMC will monitor the safety of the patients on a periodic basis. The DMC will provide a recommendation to Merrimack based on periodic safety review and/or the interim analysis whether to terminate the trial early due to safety or futility. Additionally the DMC will make a data recommendation about early termination due to observed results. In an attempt to minimize bias the interim analysis will be generated by an unblinded external vendor who will communicate results directly to the DMC. The DMC will discuss the efficacy recommendation with the Merrimack Steering Committee, if necessary. Details of the interim analysis and Steering Committee will be described in the DMC Charter.

## 5.2 General Considerations

All analyses discussed in this SAP will be generated for the final CSR. All datasets, documentation, and the SAP will be archived in the Trial Master File (TMF). Tables, figures, and listings (TFLs), as described in this document and in the mock shells, will be created using SAS version 9.2 or higher. Mock table, listing, and figure shells will be developed for detailed information on the layout. The mock tables, listings, and figure shells will not be considered as part of the SAP signoff.

## 5.3 Standards

The information presented in this section describes the standard displays/mocks and analysis to be performed. Standard table displays will include columns for treatment groups called: seribantumab+fulvestrant/exemestane (seri+BB), seribantumab+fulvestrant (seri+f), seribantumab+exemestane (seri+e), placebo+fulvestrant/exemestane (pla+BB), placebo+fulvestrant (pla+f), placebo+exemestane (pla+e), and Overall. The primary comparisons are based on seri+BB versus pla+BB.

Continuous variables, as deemed appropriate, will be summarized using the number of non-missing observations, mean, standard deviation, median, minimum and maximum values. Missing data values will be treated as missing in summarization. No imputation of missing data will be performed.

Categorical variables will be summarized using the number of non-missing observations, frequency and percentage out of the number of non-missing observations in each category. Percentages will be presented to one decimal place (e.g., xx.x). Percentages will not be presented when zero counts occur.

Enrolled will be defined as those patients signing informed consent and completing entry criteria.

All p-values will be presented as two-sided using a SAS format (pvalue5.3). Primary efficacy inferential analyses will be tested against a one-sided, alpha at 0.025 (i.e., two-sided 0.05) unless stated otherwise.

### 5.3.1 Data Cut-off

A database snapshot will be determined at the time of the confirmed number of PFS events: 195 final. The snapshot will not be performed until the minimum number of PFS events have occurred. The snapshot will include all data entered and cleaned up to that particular point in the study and data will not be changed, edited, or updated on or prior to this particular date. A separate snapshot of this data will be maintained in the biostatistics and programming area. Patients that have completed the study will require investigator signature. Continuing patients will have data reviewed and approved. Data entered after the data cut-off dates will continue to be cleaned but not reported at the time of the interim or final analysis.

A final database lock will be performed upon determination of the final PFS event. Any necessary changes to previously entered data will be documented. The final database lock will include all data for all timepoints and will be conducted at a later timeframe. Site investigators will provide signatures for any remaining patient data which was outstanding at the time of soft lock.

### 5.3.2 Analysis of Key Time-to-Event Endpoints

Time-to-event endpoints between two treatment arms will be compared with a 1-sided, $\alpha = 0.025$, stratified log-rank test, unless specified otherwise. Stratification factors will include: prior chemotherapy use (yes, no), chemotherapy backbone (exemestane or fulvestrant), and geographic region (US, Non-US/Non-Asia, Asia). Stratified Cox proportional hazards models will be constructed with stratification factors and treatment in the model to estimate hazard ratio and 2-sided 95% confidence interval (CI).

Additional comparisons between treatment groups will be performed using an unstratified log-rank test. Cox proportional hazards model with treatment group only will be constructed to estimate the hazard ratio and 2-sided 95% confidence intervals.

Potential influences of other baseline patient characteristics on various populations may be explored. Candidate covariates for the multivariate model will be selected using a stepwise selection process with treatment always in the model. Only covariates significant at a 15% level will be entered in the model and covariates significant at

15% will be considered for the final multivariate model. Treatment-by-covariate effects will be explored only for the set of factors included in the final model. The estimated hazard ratio and 2-sided 95% CI will be reported. Forest plots will be displayed for subgroup analyses.

The assumption of proportional hazards will be explored using a variety of graphical displays and testing. Hazard ratios and log (-log) of the survival probability estimates over time will be graphed. A scatterplot of smoothed Schoenfeld residuals over log of time will be displayed to check the primary efficacy analysis assumptions. The assumption of proportional hazards will be reviewed for the primary efficacy endpoint using the available graphical information. No formal testing will be performed in place of the stratified log-rank, if the assumption of proportional hazards does not hold.

Kaplan-Meier methods will be used in estimating the probability of event and displayed graphically. Time-to-event calculations will be performed in months unless specified otherwise. Median event times and 2-sided 95% CI will be provided based on Brookmeyer-Crowley methods. Probability of event estimates at 6, 9, 12, 15, and 18 month intervals will be reported in tables.

### 5.3.3 Analysis of Key Binary Endpoints

Frequencies and percentages will be reported for binary endpoints. The rate of binary endpoints between 2 treatment arms will be compared using a Cochran Mantel-Haenszel (CMH) test stratified according to the stratification factors and a Pearson $\chi^2$ test, both at a 2-sided, $\alpha = 0.05$ level. The relative risk ratio estimate will be used to contrast treatment effects on response rates. Point estimates of the rates for each treatment arm will be provided along with the corresponding exact 2-sided 95% CI using exact methods based on the binomial distribution.

### 5.3.4 Analysis of Continuous Endpoints

Mean, standard deviation, median, minimum, maximum will be reported for continuous endpoints. Least-squares mean and standard error with 95% confidence intervals will be reported for models. Models will be tested using a 2-sided, $\alpha = 0.05$ level.

## 5.4 Baseline Analysis

### 5.4.1 Disposition, Populations, and Patients Excluded from Population

Treatment termination and study termination will be summarized overall and by reason. Included in the summarization will be deaths. Overall number of deaths, deaths on-treatment (up to 30 days post last dose date), deaths after treatment.

A listing of patients and those excluded from a population along with reason will be provided, if available.

Protocol deviations will be listed and major protocol deviations will be summarized.

### 5.4.2 Demographics, Cancer Diagnosis, Medical History

Descriptive statistics will be reported for Age (years), sex (male, female), race (white/Caucasian, Black or African American, Alaska native or American Indian, Asian/Oriental, Native Hawaiian or Pacific Islander, other), height (cm), weight (kg), ethnicity, and ECOG at baseline. Other cancer diagnosis factors such as frequency of adenocarcinoma, squamous cell; TNM Stage; Bone disease, brain metastisis, visceral metastisis, number of prior chemotherapies, EGFR mutation status, KRAS status, best response will be reported. Time from cytological or histo-pathologic diagnosis to randomization will be analyzed as a continuous parameter.

Stratification factors will be summarized by frequencies and percentages. Frequency of patients randomized within each site by country will be summarized.

Prior surgical therapy (yes/no), prior radiation therapy (yes/no) will be characterized.

Medical history will be coded using the latest version of MedDRA coding. Medical history will be summarized by System Organ Class and preferred term.

### 5.4.3 Extent of Disease

Number of patients with target lesions only, non-target lesions only, and both target and non-target lesions will be summarized. Individual disease sites for target or non-target lesions will be summarized.

### 5.4.4 Prior Systemic Cancer Treatment

The following will be summarized categorically unless stated as a continuous parameter.

- Type of prior anti-cancer treatment
- Number of prior anti-cancer treatment
- Time since last anti-cancer treatment prior to study enrollment will be summarized as a continuous parameter

## 5.5 Efficacy Analysis

### 5.5.1 Primary Efficacy PFS Analysis

PFS analysis will be performed using a stratified log-rank test. Stratification factors will include: prior chemotherapy use (yes, no), chemotherapy backbone (exemestane or fulvestrant), and geographic region (US, Non-US/Non-Asia, Asia). Patients with missing stratification factor will be included in the primary efficacy analysis by merging them into the largest category for that factor. Misclassification of stratification factor will be analyzed as the actual (or ITT) analysis. A sensitivity analysis using the planned (or appropriate strata) will be performed. Seri+BB will be compared relative to pla+BB using a one-sided significance level at 0.025.

### 5.5.2 PFS Sensitivity and Subgroup Analysis

PFS will be performed using a variety of methodology to assess the sensitivity to deviations and covariate factors. The same one-sided significance level at 0.025 will be used for all situations.

- Unstratified log-rank test will be performed and a Cox proportional hazards model with treatment only in the model to estimate the hazard ratio and 95% confidence intervals
- Treatment group comparison using interval-censored methods
- Evaluation of influences of baseline covariates and biomarker factors using stepwise selection methods
- Treat all censors on the seribantumab arm as events on the date of censorship
- Cox proportional hazard models will include treatment and covariate in the model. Additional models will include treatment, covariate, and the interaction effect in the model. Hazard ratios and 95% confidence intervals will be reported for all models. In addition, median event estimates and 95% confidence intervals will be reported. Forest plots of the hazard ratio estimates and confidence intervals will be displayed for the aforementioned models.

PFS will also be analyzed using IRR. PFS will be analyzed using a stratified Cox proportional hazard models similar to the primary analysis. Cox proportional hazard models will include treatment and covariate in the

model. Additional models will include treatment, covariate, and the interaction effect in the model. Hazard ratios and 95% confidence intervals will be reported for all models. In addition, median event estimates and 95% confidence intervals will be reported. Forest plots of the hazard ratio estimates and confidence intervals will be displayed with for the aforementioned models.

### 5.5.3 OS, TTP and ORR Secondary Efficacy Analysis

OS and TTP will be analyzed in the all study populations defined earlier as time-to-event analyses. Stratifed and unstratified log-rank tests will be performed for the ITT study population.

Frequencies and percentages of ORR responders will be summarized. CMH and Pearson's chi-square test will be performed for responders. A relative risk estimator will be used to contrast the treatment effects. Risk difference estimate and 95% confidence intervals will be calculated using the normal approximation.

Waterfall plots for best response and best percentage change will be displayed for study and subpopulations.

### 5.5.4 Exploratory Biomarker Analysis

Descriptive statistics will be used to characterize the biomarkers within treatment groups over time. Continuous markers will be summarized using the change from baseline and observed values. Change from baseline to last observed will be analyzed using an ANCOVA model with baseline and stratification factors, treatment as fixed effects. Interaction effects may be tested and included in the final model(s) if deemed important. Least squares means and 95% confidence intervals will be reported. Categorical markers will be summarized using shift from baseline to last observed. CMH test using stratification factors will be used to analyze parameters.

The biomarker relationship to outcomes OS and PFS will be analyzed using a stratified log-rank test. Screening values of biomarker values will be used in the following analysis. Continuous biomarker values will be dichotomized by ≥ median, < median; as well as quartiles. Biomarker analyses will consist of performing analysis on the overall PFS endpoint. Kaplan-Meier displays will be reported. Kaplan-Meier curves will include the following:

- seribantumab, Control – According to categorical grouping
- Control Only – According to categorical grouping

Included in the graphs will be hazard ratio and 95% confidence intervals, median PFS and 95% confidence intervals, number and percentage of events and censors. These will be performed for all markers identified. Forest plots will be used to describe the biomarker subgroups relationship via hazard ratios and 95% confidence intervals.

## 5.6 Safety Analysis

All safety analysis will be performed using the safety population.

### 5.6.1 Study Drug Exposure

Extent of treatment exposure will be summarized for seribantumab, fulvestrant, and exemestane as follows:

- Number and percentage of patients beginning 1, 2, 3, …., 6 and > 6 cycles
- Number of cycles started will be treated as a continuous variable
- Actual Dose received defined as total actual dose (mg) divided by BSA ($m^2$)

- Relative Actual Dose = [Total actual dose (mg) divided by Total intended dose (mg)] x 100, where total intended dose is defined as the dose as prescribed which is summed from first to last date irrespective of missing, interruption, or reduction.
- Total Exposure (weeks) derived as (last dose date – first dose date + 1)
- Dose intensity (mg/m$^2$/wk) will be summarized as Actual Dose received divided by (duration in days multiplied by 7)

Treatment Delays and Dose Modifications

Definitions:

A **dose reduction** is defined as actual dose taken is ≤33% than the prescribed dose on any given day for any reason with the exception of a day with total dose administered of 0mg is not considered a dose reduction.

A **dose interruptions/delay/missed dose** is defined as a planned dosing day with 0 mg administered

Patients will be summarized by frequency and percentage according to:

- Patients with at least one dose reduction overall and by cycle
- Patients with at least one dose reduction due to an adverse event
- Total number of dose reductions per patient
- Number of dose interruptions/missed overall and by cycle

### 5.6.2 Adverse Events

Adverse events will be coded using the latest MedDRA dictionary at the time of coding. NCI CTCAE v4.0 will be used to grade severity. Frequencies and percentages will be reported by System Organ Class (SOC) and preferred term (PT). Patients with multiple events for a category will be counted once in the highest CTCAE classification. Summary tables will include treatment-emergent AEs, unless specified otherwise. Treatment-emergent adverse events (TEAEs) will be defined as patients based on the event onset date relative to first dose of study drug in the study. Missing onset date for an event will be considered as TEAE. Adverse events with a starting date 30 or more days after the last dose date are not considered TEAEs. Relatedness will be considered: Possibly, Probably, or Definitely related per case report form classification. If relatedness is missing the adverse event will be considered TEAE. The order of reporting the events in summary tables will be according to the following where applicable:

- SOC will be alphabetically reported
- Descending order of frequencies according to the number of patients in the seribantumab treatment group experiencing an adverse event
- PT will be alphabetically reported

The following tables will be reported:

- TEAE
  - Any TEAE by SOC
  - Any TEAE by SOC and PT*
  - Any TEAE by PT
  - Any TEAE by PT by NCI CTCAE Grade
  - Any TEAE by PT by cycle

- Any TEAE by PT by NCI CTCAE Grade by cycle

- Related TEAE
  - Any TEAE by SOC and PT*
  - Any TEAE by PT
  - Any TEAE by PT by NCI CTCAE Grade

- TEAE - NCI CTCAE Grade 3 or higher
  - Any TEAE by SOC and PT
  - Any TEAE by PT

- Related TEAE - NCI CTCAE Grade 3 or higher
  - Any TEAE by SOC and PT
  - Any TEAE by PT

- SAE – TEAEs only
  - Any SAE by SOC and PT*
  - Any SAE by PT

- TEAE - Various
  - Any TEAE Resulting in Dose Reduction
  - Any TEAE Resulting in Treatment Discontinuation
  - Any TEAE Resulting in Death

In the AE preferred term summary tables, a patient having multiple events within a System Organ Class is counted only once for that SOC. Similarly, a patient having multiple events under a preferred term is counted only once for that PT. Listings will be provided for all adverse event data.

### 5.6.3 Laboratory Evaluation

Laboratory analytes will be summarized according to panel (i.e., hematology, chemistry, urinalysis).

Shift in laboratory analyte grade will be characterized by worst from baseline and at end of treatment. Shift categories will include normal grade 1 to grade 5 and totals. Percentages will be based on the number with baseline and postbaseline values for that particular analyte.

### 5.6.4 Vital Signs

Vital sign parameters: systolic and diastolic blood pressure (mmHg), respiratory rate (breaths per minute), body temperature (C°), weight (kg) and pulse rate (beats per minute) will be summarized by observed and change from baseline according to minimum and maximum values. Additionally the frequency and percentage of patients for each observed parameter outside the normal range during treatment (up to 30 days post last dose date) will be reported.

**Table 5.6.4.1 Vital Signs Normal Range**

| Parameter | Units | Normal Range |
|---|---|---|
| Systolic Blood Pressure | mmHg | 90 - 180 |
| Diastolic Blood Pressure | mmHg | 50 - 105 |
| Pulse Rate | Beats per minute | 50 - 120 |
| Respiratory Rate | Breaths per minute | 12 – 25 |

### 5.6.5 Electrocardiogram

The number and percentage of patients with worst ECG result as order: Abnormal, Clinically Significant, Abnormal, not Clinically Significant, Normal will be summarized at screening. Results will be counted once for each patient and scheduled visit.

Exposure-Response Analysis: The relationship between seribantumab plasma concentration and ΔQTcF (change-from-baseline QTcF) will be quantified using a linear mixed effects model with ΔQTcF as the dependent variable, plasma concentration as continuous covariate, and treatment (active or placebo) and time as categorical factors, and a random intercept and slope per subject. The geometric mean of the individual Cmax values for subjects on seribantumab will be determined. The predicted effect and its two-sided 90% CIs for ΔΔQTcF (placebo-corrected change-from-baseline QTcF) at this geometric mean Cmax will be obtained.

By-Timepoint Analysis: The analysis for QTcF will be based on a linear mixed effects model with ΔQTcF as the dependent variable, baseline QTcF as covariate, and time (categorical), treatment (seribantumab and placebo), and time by treatment interaction as factors. An unstructured covariance matrix will be specified for the repeated measures at post-dose timepoints within subject. If the model with unstructured covariance matrix fails to converge, other covariance matrix such as compound symmetry can be considered. For this analysis a 90% 2-sided confidence interval (CI) will be calculated for the contrasts " seribantumab - placebo" for each time point, separately. For change-from-baseline as well as the placebo-corrected change-from-baseline HR, PR, and QRS, descriptive analysis including number of subjects, mean, median, min, max, standard deviation (SD), standard error (SE), and 90% confidence interval (CI; from t-distribution) will be calculated for each treatment and timepoint.

Categorical Analysis: The analysis results for categorical outliers and T-wave morphology will be summarized in frequency tables with counts and percentages for both number of subjects and number of timepoints. For categorical outliers, the number (percentage) of subjects as well as timepoints with increases in QTcF from baseline of >30 and >60 msec and absolute QTcF values >450, >480, and >500 msec; increase in PR interval from baseline >25% and a PR interval >200 msec; increase in QRS interval from baseline >25% and a QRS interval >120 msec; decrease in HR from baseline >25% and a HR <50 bpm; and increase in HR from baseline >25% and a HR >100 bpm will be determined by each treatment group, respectively. For T-wave morphology, the analysis will be focused on the treatment-emergent changes.

### 5.6.8 Pharmacokinetic (PK)

A Non-compartmental Analysis (NCA) will be used to analyze the pharmacokinetic parameters based on plasma concentration versus time. Mean concentration curves will be displayed using linear and semi-log scales. Standard errors will be displayed on the graphs for each collection day. A summary of PK parameters will be developed for each study drug and will include the previously indicated parameters. Additional analyses may be performed if deemed necessary.

# 6. REFERENCES

Adapted from E.A. Eisenhauer, et al: New response evaluation criteria in solid tumours:
Revised RECIST guideline (version 1.1). European Journal of Cancer 45 (2009) 228–247

Cui L, Hung HMJ, and Wang S (1999). Modification of sample size in group sequential clinical trials. Biometrics, 55, 853-857

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, *75*(4), 800–2.

# 7. APPENDICES

## CATEGORIZING LESIONS AT BASELINE

Measurable Lesions - Defined as lesions accurately measured in at least one dimension.

- Lesions with longest diameter twice the slice thickness and at least 10 mm or greater when assessed by CT or MRI (slice thickness 5-8 mm)
- Lesions with longest diameter at least 20 mm when assessed by Chest X-ray
- Superficial lesions with longest diameter 10 mm or greater when assessed by caliper
- Malignant lymph nodes with the short axis 15 mm or greater when assessed by CT.

**NOTE: Shortest axis is used as the diameter for malignant lymph nodes, longest axis for all other measurable lesions.**

Non-measurable disease - Defined as lesions too small to be considered measurable (including nodes with short axis between 10 and 14.9 mm) and truly non-measurable disease such as pleural or pericardial effusions, ascites, inflammatory breast disease, leptomeningeal disease, lymphangitic involvement of skin or lung, clinical lesions that cannot be accurately measured with calipers, abdominal masses identified by physical exam that are not measurable by reproducible imaging techniques.

- Bone disease**:** Not measurable with the exception of soft tissue components that can be evaluated by CT or MRI and meet the definition of measurability at baseline.
- Previous local treatment: A previously irradiated lesion (or lesion subjected to other local treatment) is non-measurable unless it has progressed since completion of treatment.

Normal sites

- Cystic lesions**:** Simple cysts should not be considered as malignant lesions and should not be recorded either as target or non-target disease. Cystic lesions thought to represent cystic metastases can be measurable lesions, if they meet the specific definition above. If non-cystic lesions are also present, these are preferred as target lesions.

- Normal nodes**:** Nodes with short axis <10 mm are considered normal and should not be recorded or followed either as measurable or non-measurable disease.

## RECORDING TUMOR ASSESSMENTS

All sites of disease must be assessed at baseline. Baseline assessments should be done as close as possible prior to study start. For an adequate baseline assessment, all required scans must be done within 28 days prior to treatment and all disease must be documented appropriately. If baseline assessment is inadequate, subsequent statuses generally should be not evaluable.

***Target lesions***

- All measurable lesions up to a maximum of 2 lesions per organ, 5 lesions in total, representative of all involved organs, should be identified as target lesions at baseline. Target lesions should be selected on the basis of size (longest lesions) and suitability for accurate repeated measurements. Record the longest diameter for each lesion, except in the case of pathological lymph nodes for which the short axis should be recorded. The sum of the diameters (longest for non-nodal lesions, short axis for nodal lesions) for all target lesions at baseline will be the basis for comparison to assessments performed on study.
- If two target lesions coalesce the measurement of the coalesced mass is used. If a large target lesion splits, the sum of the parts is used.
- Measurements for target lesions that become small should continue to be recorded. If a target lesion becomes too small to measure, 0 mm should be recorded if the lesion is considered to have disappeared; otherwise a default value of 5 mm should be recorded.

**NOTE: When nodal lesions decrease to <10 mm (normal), the actual measurement should still be recorded.**

***Non-target disease***

All non-measurable disease is non-target. All measurable lesions not identified as target lesions are also included as non-target disease. Measurements are not required but rather assessments will be expressed as ABSENT, UNEVALUATED, PRESENT, UNEQUIVOCAL PROGRESSION. Multiple non-target lesions in one organ may be recorded as a single item on the case report form (e.g., 'multiple enlarged pelvic lymph nodes' or 'multiple liver metastases').

## OBJECTIVE RESPONSE STATUS AT EACH EVALUATION

Disease sites must be assessed using the same technique as baseline, including consistent administration of contrast and timing of scanning. If a change needs to be made the case must be discussed with the radiologist to determine if substitution is possible. If not, subsequent objective statuses are not evaluable.

***Target disease***

- Complete Response (CR): Complete disappearance of all target lesions with the exception of nodal disease. All target nodes must decrease to normal size (short axis < 10 mm). All target lesions must be assessed.
- Partial Response (PR): Greater than or equal to 30% decrease under baseline of the sum of diameters of all target measurable lesions. The short diameter is used in the sum for target nodes, while the longest diameter is used in the sum for all other target lesions. All target lesions must be assessed.
- Stable: Does not qualify for CR, PR or Progression. All target lesions must be assessed. Stable can follow PR only in the rare case that the sum increases by less than 20% from the nadir, but enough that a previously documented 30% decrease no longer holds.
- Documented Progression Disease (PD): 20% increase in the sum of diameters of target measurable lesions above the smallest sum observed (over baseline if no decrease in the sum is observed during therapy), with a minimum absolute increase of 5 mm.
- Not Evaluable: Progression has not been documented, and
  - one or more measurable target lesions have not been assessed
  - or assessment methods used were inconsistent with those used at baseline
  - or one or more target lesions cannot be measured accurately (e.g., poorly visible unless due to being too small to measure)
  - or one or more target lesions were excised or irradiated and have not reappeared or increased

***Non-target disease***

- CR: Disappearance of all non-target lesions and normalization of tumor marker levels. All lymph nodes must be 'normal' in size (<10 mm short axis).
- Non-CR/Non-PD: Persistence of any non-target lesions and/or tumor marker level above the normal limits.
- PD: Unequivocal progression of pre-existing lesions. Generally the overall tumor burden must increase sufficiently to merit discontinuation of therapy. In the presence of SD or PR in target disease, progression due to unequivocal increase in non-target disease should be rare.
- Not Evaluable: Progression has not been determined and one or more non-target sites were not assessed or assessment methods were inconsistent with those used at baseline.

***New Lesions***

The appearance of any new unequivocal malignant lesion indicates PD. If a new lesion is equivocal, for example due to its small size, continued assessment will clarify the etiology. If repeat assessments confirm the lesion, then progression should be recorded on the date of the initial assessment. A lesion identified in an area not previously scanned will be considered a new lesion.

***Supplemental Investigations***

If CR determination depends on a residual lesion that decreased in size but did not disappear completely, it is recommended the residual lesion be investigated with biopsy or fine needle aspirate. If no disease is identified, objective status is CR.

If progression determination depends on a lesion with an increase possibly due to necrosis, the lesion may be investigated with biopsy or fine needle aspirate to clarify status.

***Symptomatic/clinical progression***

Subjects requiring discontinuation of treatment without objective evidence of disease progression should not be reported as PD on tumor assessment CRFs. Every effort should be made to document objective progression even after discontinuation of treatment.

**Table 1. Objective Response Status at each Evaluation**

| Target Lesions | Non-target Lesions | New Lesions | Objective Status |
|---|---|---|---|
| CR | CR | No | CR |
| CR | Non-CR/Non-PD | No | PR |
| CR | NE, or Missing | No | PR |
| PR | Non-CR/Non-PD, NE, or Missing | No | PR |
| SD | Non-CR/Non-PD, NE, or Missing | No | SD |
| NE or Missing | Non-PD | No | NE |
| PD | Any | Yes or No | PD |
| Any | PD | Yes or No | PD |
| Any | Any | Yes | PD |