Boston Biomedical, Inc.

BBI608-336

A Phase III Randomized, Double-Blind, Placebo-Controlled Clinical Trial of BBI608 plus Weekly Paclitaxel vs. Placebo plus Weekly Paclitaxel in Adult Patients with Advanced, Previously Treated Gastric and Gastro-Esophageal Junction Adenocarcinoma

Statistical Analysis Plan

[redacted]

Version 2.0
Date: 27 March 2017

SIGNATURE PAGE

Prepared by:

[redacted]

Reviewed by:

[redacted]

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AE	adverse event
ALT	alanine transaminase
AST	aspartate transaminase
BBI	Boston Biomedical, Inc.
BOR	best overall response
BP	blood pressure
bpm	beats per minutes
CBC	complete blood count
CMH	Cochran-Mantel-Haenszel
CI	confidence interval
CR	complete response
CSC	cancer stem cell
CTCAE	common toxicity criteria for adverse events
DCR	disease control rate
dL	deciliter
DSMB	Data Safety and Monitoring Board
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	electronic case report form
EDC	electronic data capture
EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 Item
FDA	(United States) Food and Drug Administration
FFPE	Formalin Fixed Paraffin Embedded
GEJ	Gastro-Esophageal Junction
HR	hazard ratio
ICF	informed consent form
ICH	International Conference on Harmonization
IDE	investigational device exemption
IHC	immunohistochemical
IRB	Institutional review Board
ITT	intent-to-treat
IV	intravenous
IxRS	Interactive Web/Voice Response System
LDH	lactate dehydrogenase
LLN	lower limit of normal
LNH	low-normal-high
mckat	microkats
MedDRA	Medical Dictionary for Regulatory Activities
MET	minimum effective treatment
mg/m	milligrams per minute
mmHg	millimeter Mercury

mmol	millimole
NCI CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
ORR	objective response rate
OS	overall survival
PD	progressive disease
PFS	progression-free survival
PK	pharmacokinetic
PP	per protocol
PR	partial response
PS	performance status
PT	preferred term
PTH	Plasma Parathyroid Hormone
QoL	Quality of Life
QRS	portion of the electrocardiogram comprising the Q, R, and S waves, together representing ventricular depolarization
QT	Q wave to T wave interval
QTcB	QT Interval Bazett Correction
QTcF	QT Interval Fridericia Correction
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	serious adverse event
SAP	statistical analysis plan
SAS	Statistical Analysis System
SD	stable disease or standard deviation
SGOT	Serum Glutamic Oxaloacetic Transaminase
SGPT	Serum Glutamic Pyruvate Transaminase
SMQ	Standard MedDRA queries
SOC	system organ class
U/L	unit per liter
ULN	upper limit of normal
µmol	micromole

1 INTRODUCTION

Gastric cancer and cancer of the gastro-esophageal junction (GEJ) are a significant cause of morbidity and mortality worldwide and have the second highest death rate of all cancers with 736,000 deaths recorded in 2008. Despite poor survival rates in gastric and GEJ cancer, the past decade has seen some progress in second-line therapy for this disease with approximately 20-50% of subjects receiving second-line therapy world-wide. Second-line therapy has shown a median survival of 5.6 months as compared to survival of approximately 2.5 months in subjects not receiving treatment following progression on first-line therapy.

BBI608 is the most advanced agent developed by Boston Biomedical, Inc. (BBI) to target cancer stem cells (CSCs). BBI608 is a small molecule that blocks self-renewal of, and induces cell death in CSCs isolated from gastric and GEJ cancer as well as other types of cancer. The unmet clinical need, mechanism of action of BBI608, and encouraging pre-clinical and clinical data seen to date provide a strong rationale for further investigation of BBI608 in combination with weekly paclitaxel in subjects with advanced gastric and GEJ cancer.

This Statistical Analysis Plan (SAP) is created based on BBI608-336 Protocol Amendment 7 dated December 2nd 2016, EudraCT Number 2014-000774-18 (Mar 7, 2014), and it describes in detail the statistical methodology and the statistical analyses to be conducted for the above mentioned protocol. The analysis plan may change due to unforeseen circumstances. Any deviations to the planned statistical analyses specified within the SAP will be justified in writing and presented within the clinical study report.

2 STUDY OBJECTIVES

2.1 Primary Objective

- To compare Overall Survival (OS), defined as the time from randomization until death from any cause, in subjects with pre-treated advanced gastric/GEJ adenocarcinoma treated with BBI608 plus weekly paclitaxel *versus* placebo plus weekly paclitaxel in the intent-to-treat (ITT) general study population.

2.2 Secondary Objectives

- To compare OS, defined as time from randomization until death from any cause, in subjects with pre-treated advanced gastric/GEJ adenocarcinoma treated with BBI608 plus weekly paclitaxel *versus* placebo plus weekly paclitaxel in the predefined biomarker-positive sub-population.
- To compare Progression-Free Survival (PFS), defined as the time from randomization until the first objective observation of disease progression or death from any cause, in subjects with pre-treated advanced gastric/GEJ adenocarcinoma treated with BBI608 plus weekly paclitaxel *versus* placebo plus weekly paclitaxel in the ITT general study population.

- To compare PFS, defined as the time from randomization until the first objective observation of disease progression or death from any cause, in subjects with pre-treated advanced gastric/GEJ adenocarcinoma treated with BBI608 plus weekly paclitaxel *versus* placebo plus weekly paclitaxel in the predefined biomarker-positive subpopulation.
- To compare the Objective Response Rate (ORR), defined as the proportion of subjects with a documented complete response or partial response (CR + PR) based on Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 criteria, in subjects with pre-treated advanced gastric/GEJ adenocarcinoma treated with BBI608 plus weekly paclitaxel *versus* placebo plus weekly paclitaxel in the ITT general study population.
- To compare the Disease Control Rate (DCR), defined as the proportion of subjects with a documented complete response, partial response or stable disease (CR + PR + SD) based on RECIST 1.1 criteria, in subjects with pre-treated advanced gastric/GEJ adenocarcinoma treated with BBI608 plus weekly paclitaxel *versus* placebo plus weekly paclitaxel in the ITT general study population.
- To evaluate the safety profile of BBI608 administered daily plus weekly paclitaxel in subjects with pre-treated advanced gastric/GEJ adenocarcinoma, with safety assessed according to the National Cancer Institute Common Toxicity Criteria for Adverse Events (NCI CTCAE version 4.0).

2.3 Correlative Objectives

- To explore the exposure/response relationships of BBI608 and paclitaxel in subjects with pre-treated advanced gastric/GEJ adenocarcinoma treated with BBI608 plus weekly paclitaxel *versus* placebo plus weekly paclitaxel using population pharmacokinetics with sparse PK sample collection.
- To explore an association between putative biomarkers as determined from paraffin-embedded tumor specimens and the potential for clinical benefit in terms of overall survival, progression-free survival, disease control rate, and objective response rate, from treatment with BBI608 plus weekly paclitaxel in subjects with pre-treated advanced gastric/GEJ adenocarcinoma.
- To explore associations with baseline values and changes of putative biomarkers in the blood and the potential for clinical benefit in terms of overall survival, progression-free survival, disease control rate, and objective response rate, from treatment with BBI608 plus weekly paclitaxel in subjects with pre-treated advanced gastric/GEJ adenocarcinoma.

2.4 Exploratory Objectives

- To compare the Objective Response Rate (ORR), defined as the proportion of subjects with a documented complete response or partial response (CR + PR) based on RECIST 1.1 criteria, in subjects with pre-treated advanced gastric/GEJ adenocarcinoma treated with BBI608 plus weekly paclitaxel *versus* placebo plus weekly paclitaxel in the predefined biomarker-positive subpopulation.

- To compare the Disease Control Rate (DCR), defined as the proportion of subjects with a documented complete response, partial response or stable disease (CR + PR + SD) based on RECIST 1.1 criteria, in subjects with pre-treated advanced gastric/GEJ adenocarcinoma treated with BBI608 plus weekly paclitaxel *versus* placebo plus weekly paclitaxel in the predefined biomarker-positive subpopulation.
- To characterize the population pharmacokinetics of BBI608 in subjects with pre-treated advanced gastric/GEJ adenocarcinoma, examining how demographic and pathophysiologic factors affect plasma concentration of BBI608. The exposure-response relationship between clinical and safety endpoints and BBI608/paclitaxel exposure will also be examined.
- To compare the Quality of Life (QoL), as measured using the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 Item (EORTC QLQ-C30), in subjects with pre-treated advanced gastric/GEJ adenocarcinoma treated with BBI608 plus weekly paclitaxel *versus* placebo plus weekly paclitaxel in the ITT general study population.

3 INVESTIGATIONAL PLAN

3.1 Overall Study Design and Plan

This is a randomized, double-blind, multi-center, phase III study of BBI608 plus weekly paclitaxel vs. placebo plus weekly paclitaxel for adult subjects with advanced, pre-treated gastric or GEJ adenocarcinoma who have failed first line therapy containing a fluoropyrimidine and a platinum-based agent.

A total of 700 subjects will be randomized in a 1:1 ratio, stratified according to prior taxane therapy (yes vs. no); disease measurability by RECIST 1.1 (measurable disease present vs. not present); time to progression on first line therapy (<6 months vs. ≥6 months); and geographical region (Asia vs. North America, Europe and Australia vs. South America). These subjects will be randomized over a 24 month period, and followed for an additional 12 months.

The study will proceed in 28-day (4-week) cycles. BBI608 or placebo will be administered orally, twice daily, with doses separated by 12 hours. Paclitaxel 80 mg/m^2 IV will be administered weekly, on day 1, 8, and 15 of each 28 day study cycle. BBI608 or placebo administration will begin 2 days prior to the first paclitaxel infusion.

Tumor assessments will be performed every 8 weeks after randomization until objective disease progression or another discontinuation criterion is met.

This study will enroll subjects with a cytologically or histologically confirmed adenocarcinoma of the stomach or gastro-esophageal junction (GEJ) that is metastatic or locally advanced and unresectable. Subjects will have failed treatment with one regimen containing at least a platinum/fluoropyrimidine doublet for unresectable or metastatic disease.

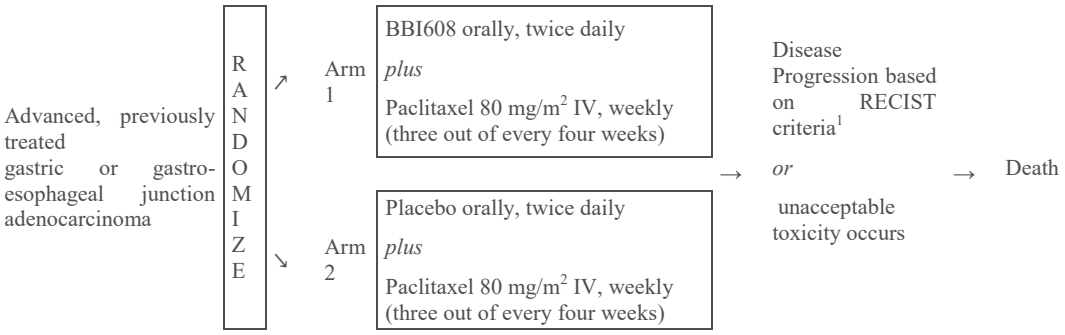
Subjects will have received no more than one regimen in the metastatic setting. Subjects who have

received prior taxane therapy may be enrolled, so long as prior taxane was administered in the adjuvant or neoadjuvant setting and progression occurred >6 months of end of treatment. Subjects who have received prior taxane in first line metastatic setting are excluded.

Other inclusion criteria for all subjects include: age ≥ 18 yrs.; ECOG performance status ≤ 1; and adequate hepatic, renal, and bone-marrow function. Detailed study inclusion and exclusion criteria are provided in the protocol.

Subjects in this study will receive BBI608 or matched placebo orally, daily, at 480 mg bid (960 mg total daily dose). In each cycle BBI608 will be taken twice daily for 4 weeks (28 days). BBI608 or placebo will be administered twice daily, one hour prior or two hours after meals, with the first dose taken in the morning and doses separated by approximately 12 hours.

Subjects will receive BBI608 or placebo in combination with paclitaxel, 80 mg/m^{2}administered via 1-hour IV infusion on days 1, 8, and 15 of each 28 day cycle. Dose modification of BBI608/placebo and/or paclitaxel is allowed.



[1]If no other standard therapies are available at the time of disease progression, and the subject has not experienced any adverse events requiring permanent discontinuation, BBI608/Placebo may be continued in monotherapy as long as it is the opinion of the Investigator that the subject may continue to be deriving benefit.

Subjects may continue to receive protocol therapy as long as they have not experienced any adverse events requiring permanent discontinuation of study medication and have not demonstrated disease progression based on RECIST criteria. If no other standard therapies are available at the time of disease progression, and the subject has not experienced any adverse events requiring permanent discontinuation, BBI608/Placebo may be continued in monotherapy as long as it is the opinion of the Investigator that the subject may continue to be deriving benefit. If either BBI608/Placebo or paclitaxel is discontinued due to toxicity, the other agent may be continued as monotherapy until another discontinuation criterion is met.

There will be one interim analysis in this study to be presented to the independent Data Safety and Monitoring Board (DSMB). Additionally, DSMB will review safety during the entire conduct of the study. The role and responsibility of the DSMB will be defined in a separate Charter (see SAP section 4.10.9).

The interim analysis will be performed on OS, the primary endpoint of this study, when at least 2/3^{rds} of the required number of events (380) has been observed in the ITT population.

3.2 Efficacy and Safety Variables

3.2.1 Efficacy Variables:

3.2.1.1 Primary Efficacy Endpoints:

- Overall Survival in the ITT general study population

3.2.1.2 Secondary Efficacy Endpoints:

- Overall Survival in the predefined biomarker-positive sub-population
- Progression-Free Survival in the ITT general study population
- Progression-Free Survival in the predefined biomarker-positive sub-population
- Objective Response Rate in the ITT general study population
- Disease Control Rate in the ITT general study population

3.2.2 Safety Endpoints

- Incidence of treatment emergent adverse events (TEAEs), treatment emergent events of special interest (AESI), and serious adverse events (SAEs)
- Incidence of laboratory toxicities
- Physical examination findings
- Observed and changes in standard clinical laboratory parameters (hematology, biochemistry)
- Observed and changes in vital sign measurements
- Observed and changes in electrocardiogram (ECG) recording
- Proportion of subjects with concomitant medications and procedures
- Study drug exposure

3.2.3 Correlative Endpoints

The correlative science component of the BBI608-336 trial will include tumor and blood-based assays to identify biomarkers of benefit from BBI608 therapy, as well as biomarkers of BBI608 resistance. The purpose of these studies is to explore the relationship between biomarkers and disease response, as measured by following endpoints:

- Overall survival
- Progression-Free Survival
- Objective Response Rate
- Disease Control Rate

3.2.4 Exploratory Endpoints

- Objective Response Rate in the Predefined Biomarker-positive Population
- Disease Control Rate in the Predefined Biomarker-positive Population
- Sparse Pharmacokinetic Parameters
- Observed and changes in Quality of Life

4 STATISTICAL METHODS

4.1 Data Quality Assurance

All tables, figures and data listings to be included in the report will be independently checked for consistency, integrity and in accordance with standard PAREXEL procedures.

4.2 General Presentation Considerations

'Baseline' is defined as the last observation prior to first dose of study drug, including either BBI608/placebo or paclitaxel, whichever is administered first.

Continuous data will be summarized in terms of the mean, standard deviation (SD), median, minimum, maximum and number of observations, unless otherwise stated. Continuous data that are expected to be skewed will be presented in terms of the maximum, upper quartile, median, lower quartile, minimum and number of observations. The minimum and maximum will be reported to the same number of decimal places as the raw data recorded in the database. The mean, median, lower quartile and upper quartile will be reported to one more decimal place than the raw data recorded in the database. The SD will be reported to two more decimal places than the raw data recorded in the database. In general, the maximum number of decimal places reported shall be four for any summary statistic.

Categorical data will be summarized in terms of the number of subjects providing data at the relevant time point (n), frequency counts and percentages. In cases where missing data cause percentages not to sum to 100, a missing data row will be provided.Percentages will be presented to one decimal place. Percentages will not be presented for zero counts. Percentages will use column totals as the denominator unless otherwise indicated. Changes from baseline in categorical data will be summarized using shift tables where appropriate.

P-values greater than or equal to 0.0001 will be presented to four decimal places. P-values less than 0.0001 will be presented as "<0.0001".

Confidence intervals will be presented to one more decimal place than the raw data.

All report outputs will be produced using SAS® version 9.3 or later in a secure and validated environment. All report outputs will be provided to the Sponsor in a single Microsoft Word 2010 document.

4.3 General Data Handling

- Blinding Maintenance:

To ensure concealment of treatment allocation in this double-blind study, randomization will be performed using web-based access to a central database provided through the Randomization and Trial Supply Management (RTSM) system. Investigators, patients and pharmacists will be blinded to treatment assignment. Files containing treatment allocation information are created by an RTSM statistician and stored in RTSM (at PAREXEL Informatics), who is independent of trial conduct. No one outside the RTSM personnel working on the study will have access to treatment allocation information,─except the separated unblinded PAREXEL team (unblinded biostatisticians and unblinded statistical programmers) preparing unblinded outputs for review by DSMB members exclusively. The unblinded PAREXEL team will perform the unblinding according to PAREXEL SOP and has no other role in this study; in particular, unblinded PAREXEL biostatisticians were not involved in the creation of the SAP.

Additionally, as noted in the study DSMB Charter, DSMB members are unblinded and will interact with Sponsor as noted in the Charter, specifically: Sponsor will remain blinded to the trial and not have access to any unblinding information involved with the DSMB process.

- Age will be computed from July 1 of the year of birth to the date of Informed Consent, as (Date of Informed Consent - July 1 of the year of Birth +1)/365.25

- Missing Data: All analyses and descriptive summaries will be based on the observed data. Unless otherwise specified, missing data will not be imputed or "carried forward."

- Baseline Measurements:

  - Efficacy: The last measurement on or prior to the date of randomization will serve as the baseline measurement. In the event such a value is missing, the last assessment completed prior to the first study drug administration (either BBI608/placebo or paclitaxel, whichever was administered first) will be used as the baseline assessment so long as this assessment was taken within 21 days of randomization.

  - Safety: The last measurement prior to the first study drug administration (either BBI608/placebo or paclitaxel, whichever was administered first) will be used as the baseline assessment.

  - Other baseline characteristics: The last measurement on or prior to the date of randomization will serve as the baseline measurement. In the event such a value is missing, the last assessment completed prior to the first dose of study drug administration (either BBI608/placebo or paclitaxel, whichever was administered first) will be used as the baseline assessment. Please see

table below for the date cut offs accepted for each of the baseline characteristics performed prior to randomization.

Investigations		Timing prior to randomization
History and Physical Exam including:	• Prior medical and therapeutic history • Physical examination • Vital signs • Height, weight, ECOG performance status • Clinical tumor measurements • Concurrent medication list	≤14 days
Hematology	• CBC + differential • Platelet count	≤14 days[1]
Biochemistry	• Creatinine, Total Bilirubin, AST, ALT, Alkaline Phosphatase, LDH, Albumin, Potassium, Magnesium, Phosphate	≤14 days[1]
Urinalysis	• Dipstick (including protein, specific gravity, glucose and blood)	≤14 days[1]
Cardiac Assessment	• ECG (12 lead)	≤28 days
Radiology & Imaging	• CT/MRI scan of chest/abdomen/pelvis with tumor measurement and evaluation by RECIST 1.1 criteria	≤ 21 days
Correlative Studies	• Submission of representative diagnostic tumor tissue	On request
	• Blood sample collection	≤14 days
Other Investigations	• Serum or urine pregnancy test	≤ 5 days
Adverse Events[2]	• Baseline adverse event evaluation (to document residual adverse event from previous therapy and baseline symptoms)	≤14 days
Quality of Life	• EORTC QLQ-C30	≤14 days

1 If required laboratory tests cannot be performed within indicated timelines due to technical reasons, lab retest and prolongation of the screening period for 3 working days is allowed.
2 Adverse events will be graded according to the NCI Common Terminology Criteria for Adverse Events version 4.0.

- Study Day:
  - For safety analysis: Study day is calculated as:
    - Assessment date – first dose date + 1; if the assessment was performed on or after the first dose day.
    - Assessment date – first dose date; if the assessment was performed prior to the first dose date.
  - For efficacy (time to event) analysis: Study day is calculated as:
    - Assessment date – randomization date + 1; if the assessment was performed on or after the randomization date.
    - Assessment date – randomization date; if the assessment was performed prior to the randomization date.

- Unless otherwise stated, missing or partial dates will be imputed.

- Time-to-event: The event or censoring time (days) is calculated as:
  - Date of event/censoring – Date of randomization + 1

- Duration: Duration (except for duration of study treatment) is calculated as:
  - Duration (days): (End Date – Start Date + 1)

  - Duration (weeks): (End Date – Start Date + 1) / 7

  - Duration (months): (End Date – Start Date + 1) / 30.4375; (Days in months =average number of days in a year /12)

  - Duration (years): (End Date – Start Date + 1) / 365.25; (Average days in a year = 365.25, reflecting the Julian Year of three years with 365 days each and one leap year of 366 days)

4.4 Study Subjects

4.4.1 Disposition of Subjects

All subjects enrolled (signed informed consent), randomized on the study will be included in the summary of subject recruitment over time by region, site, and treatment arm. The intent-to-treat (ITT) subjects, per-protocol (PP), biomarker positive, and safety populations will be summarized by treatment arm. The frequency and percentage of subjects who discontinue before completing the study, along with discontinuation reason, will be summarized for the ITT population. The number and percentage of subjects in the ITT population who receive study treatment in each cycle will also be summarized.

Subjects who discontinue (i.e., do not complete the study per protocol) as well as subjects who are randomized but do not take a single capsule of study drug or placebo will be listed. A listing of all enrolled subjects and the study populations to which they belong will be produced.

4.4.2 Protocol Deviations

Protocol deviations (PD) and analysis sets will be reviewed and confirmed through s blind data review meeting when deciding which subjects and/or subject data will be excluded from certain analyses prior to database lock. This process will occur prior to treatment code unblinding, in a documented and approved manner. The PD definition list will be included in data review meeting minutes as an attachment.

Major protocol deviations are defined as those deviations from the protocol likely to have an impact on the perceived efficacy and/or safety of study treatments. The impact of major protocol deviations on the efficacy and/or safety results will be investigated by assessing the robustness of the study results and conclusions to the choice of analysis population, both including and excluding data potentially affected by major protocol deviations.

Major protocol deviations and any action to be taken regarding the exclusion of subjects or affected data from specific analyses are defined in the Protocol Deviation Specification, Version 2.0 (Jan 8, 2016). These deviations must be finalized prior to database lock.

Information related to minor protocol deviations, along with actions to be taken for analysis, can be found in the Protocol Deviation Specification, Version 2.0 (Jan 8, 2016).

A summary of the number of events and percentage of events divided by total events with any major protocol deviation, along with each type of major protocol deviation will be provided by treatment group and overall. A by-subject listing of major protocol deviations will be provided sorted by treatment group, site and subject.

4.5 Analysis Populations

The intent-to-treat (ITT) population will consist of all randomized subjects with treatment assignment designated according to initial randomization. The ITT population will be used for the analysis of the primary efficacy endpoint (OS), secondary efficacy endpoints (PFS, ORR, DCR), the QOL exploratory endpoint, and the correlative endpoints.

The biomarker positive sub-population of subjects in the ITT population will be used for the analysis of OS, PFS, ORR and DCR as part of secondary end point analysis.

The per-protocol (PP) population is defined as all randomized subjects who have no major protocol deviations that will confound the effects of treatment. Deviations include failure to satisfy entry criteria, failure to satisfy requirement for study treatment compliance or treatment with prohibited medications believed to confound the effects of study treatment while actively receiving study treatment. This list will be identified in a blinded fashion by sponsor review prior to database lock. Treatment assignment will be designated according to initial randomization. The PP population will be used as a sensitivity analysis for the assessment of the primary efficacy endpoint (OS).

The Minimum Effective Treatment (MET) population is defined as all randomized subjects who received at least 480 mg BBI608/placebo or higher with Paclitaxel 80 mg/m^2 weekly for at least 80% of the first 56 days of study treatment.

The safety population will include all randomized subjects who receive at least one dose of either study drug, with treatment assignment designated according to the actual study treatment received. This population will be used for all safety, as well as pharmacokinetic, analyses.

Upon database release, protocol deviation and analysis population outputs will be produced and will be sent to Sponsor for review. An analysis population classification meeting will be arranged to discuss the outputs and to decide which subjects and/or subject data will be excluded from certain analyses. Decisions made regarding the exclusion of subjects and/or subject data from analyses will be made prior to unblinding and will be documented and approved by Sponsor. Upon database lock, final analysis population outputs will be produced and will be sent to Sponsor for approval, prior to unblinding.

A summary of the number and percentage of subjects by treatment group and overall for each analysis population will be provided.

A by-subject listing of analysis population details will be provided. This listing will be presented by treatment group and should include: center, subject identifier, and inclusion/exclusion flags for each population and reason for exclusion from each population.

4.6 Demographic and Other Baseline Characteristics

The following demographic and baseline characteristics will be summarized for subjects in the ITT population by treatment arm:

- Age, age group (< 65 vs. ≥65 years), sex, race, ethnicity, ECOG status
- Weight (kg), height (cm), child-bearing potential, and body mass index (BMI)
- Randomization stratification factors: geographic region (Asia, North America/Europe/Australia, South America), time to progression on first line therapy (<6 months, ≥6 months from start of first line therapy), disease measurability by RECIST 1.1 (measurable disease present, not present), and prior taxane therapy (yes, no)

Demographic and baseline characteristics will also be listed for each subject.

4.7 Disease Characteristics, Medical History, and Prior Therapy

The following disease characteristics, medical history, and prior therapy will be summarized for subjects in the ITT population by treatment arm and overall:

- Time from initial diagnosis to randomization (months), time from most recent relapse or staging to randomization (in months), pathologic diagnosis, stage at screening, and metastatic sites of disease
- Prior radiation indication, prior chemotherapy setting, prior gastrectomy (partial or complete)
- Medical History (by SOC and PT)

Medical history and Prior therapy, surgery, or radiotherapy will be coded using the latest MedDRA dictionary version available.

Duration of gastric cancer diagnosis, and prior therapy, will be calculated in months (rounded to the nearest integer) as ((date of randomization - date of first diagnosis of gastric or GEJ adenocarcinoma/first treatment for gastric cancer) + 1) / (365.25/12). Handling of partial dates will follow the conventions in Section 4.10.1.6

Medical history will be provided in a listing. The listing will be sorted by treatment group, subject identification number, system organ class, preferred term and reported term.Medical history must include date of diagnosis including histological documentation of malignancy, documentation of *Her2* status of tumor (if available), prior anticancer therapy and prior date(s) of disease progression.

4.8 Measures of Drug Administration & Treatment Compliance

BBI608/Placebo

Several measures of drug administration will be reported. Raw treatment compliance for BBI608/Placebo is defined as the ratio, expressed as a percentage, of the number of capsules (BBI608 or placebo) taken by a subject over the course of a time interval to the number of capsules intended to be taken over that same time interval as per the dose prescribed by the Investigator, including the starting dose-level of 480 mg twice daily.

Additionally, for BBI608/placebo, daily treatment compliance will also be reported. Daily treatment compliance is defined as the ratio, expressed as a percentage, of the number of ***days*** a given patient received a defined total dose of BBI608/placebo out of total duration of treatment (days). Daily compliance will be reported for each patient for the following dose-levels and intervals:

- The % of days the patient received a total dose of 960 mg BBI608/placebo out of total duration of treatment (days) (total days dosed = 1+ [date last dose BBI608/placebo] - [date of first dose BBI608/placebo]).
- The % of days the patient received a total dose of at least 480 mg BBI608/placebo or higher out of total days dosed (total days dosed = 1+ [date last dose BBI608/placebo] - [date of first dose BBI608/placebo]).
- The % of days the patient received a non-zero dose of BBI608/placebo out of the total days dosed (total days dosed = 1+ [date last dose BBI608/placebo] - [date of first dose BBI608/placebo]).
- The % of days the patient received a total dose of 960 mg BBI608/placebo out of the first 56 days of BBI608/placebo dosing (approximately the first two study cycles)
- The % of days the patient received a total dose of at least 480 mg BBI608/placebo or higher out of the first 56 days of BBI608/placebo dosing (approximately the first two study cycles)
- The % of days the patient received a non-zero dose of BBI608/placebo out of the first 56 days of BBI608/placebo dosing (approximately the first two study cycles)

Paclitaxel

Treatment compliance for paclitaxel is defined as the ratio, expressed as a percentage, of the amount of paclitaxel administered to a subject (milligrams/m^2) over the course of a time interval to amount of paclitaxel intended to be administered over that same time interval as per the dose prescribed by the Investigator.

BBI608/placebo treatment compliance in both arms will be monitored by drug accountability with amount of BBI608/placebo consumed advised by information recorded in the subject diary. A summary of the measures of drug administration and treatment compliance by treatment group and time interval will be provided.

In addition, a patient will be considered to have received a Minimum Effective Treatment if they

received at least 480 mg BBI608/placebo or higher with Paclitaxel 80 mg/m^2 weekly for at least 80% of the first 56 days of study treatment. The number and percentage of patients who received the "Minimum Effective Treatment" will be provided in a table.

Study drug administration and compliance data will be listed for all subjects.

4.9 Efficacy Evaluation

4.9.1 Analysis and Data Conventions

The primary efficacy analysis of overall analysis will take place after the required number of events (566) has occurred.

The primary efficacy analysis will use a two-sided log rank significance test with an overall alpha of 0.05 after controlling for alpha at interim analysis (e.g., two-sided alpha= 0.012).

This study is designed to test for superiority of BBI608 plus weekly paclitaxel vs. placebo plus weekly paclitaxel. The null hypothesis states that there is no difference in overall survival between the two treatment groups. The alternative hypothesis states that BBI608 plus weekly paclitaxel is superior to placebo plus paclitaxel with longer OS. Symbolically, this is expressed as follows:

H_0: survival on BBI608 + paclitaxel = survival on placebo + paclitaxel

vs.

H_1: survival on BBI608+paclitaxel > survival on placebo + paclitaxel

Superiority will be concluded if there is sufficient evidence to reject the null hypothesis, with survival being greater in the BBI608 plus weekly paclitaxel treatment group compared to the placebo plus weekly paclitaxel treatment group.

4.9.1.1 Multi-center Studies

For the purpose of the summaries and analyses, the term 'Center' will be used to define each investigator site.

Summaries of demographic data, treatment compliance data and primary efficacy variable data by treatment group will be provided. Primary efficacy variable data by treatment group and by country will be provided. Section 4.9.1.2 provides details of adjustments for covariates such as stratification factors.

4.9.1.2 Adjustments for Covariates

The covariates listed below, which are the randomization stratification factors, will be used as adjustors in the primary and secondary efficacy analyses. They will be categorized as follows:

- Geographical region (Asia *versus* North America, Europe, and Australia *versus* South America)

- Time to progression on first line therapy (<6 months *versus* ≥6 months from start of first line therapy)
- Disease measurability by RECIST 1.1 (measurable disease present *versus* not present)
- Prior taxane therapy (yes *versus* no)

If any randomization stratum proves to have an inadequate sample size, it will be dropped from the primary analysis or pooling will be applied. Details will be developed at blinded data review meetings.

4.9.1.3 Efficacy Data Handling and Definitions

All analyses and descriptive summaries will be based on the observed data. Unless otherwise specified, missing data will not be imputed.

In addition, a Kaplan-Meier plot of the time to premature study discontinuation will be provided, identifying the recorded reason for premature discontinuation.

4.9.1.3.1 Overall Survival

The OS time is defined as the time from the date of randomization to the date of death from any cause. If a subject is not known to have died on or before the date of data cut-off, OS data will be censored on the last date (on or before the cut-off date) the subject was known to be alive. Subjects who lack data beyond randomization will have their survival times censored on the date of randomization.

4.9.1.3.2 Progression-Free Survival

PFS time is defined as the time from the date of randomization until the date of first radiographic documentation of progression as defined by RECIST (Version 1.1), or death due to any cause, whichever is first.

Imaging will be performed every 8 weeks (± 5 days) following randomization. The following table summarizes the censoring rules for the PFS analysis:

Situation	Date of Event or Censor	Event / Censor
No baseline radiological tumor assessment available	Date of randomization	Censored
No post baseline radiological tumor assessment available and no death reported within 2 scan intervals following randomization	Date of randomization	Censored
No post baseline radiological tumor assessment available but death reported within 2 scan intervals following randomization	Date of Death	Event
No tumor progression (per RECIST 1.1) and no death reported within 2 scan intervals following last adequate radiological tumor assessment	Date of last adequate radiological tumor assessment	Censored
No tumor progression (per RECIST 1.1) but death reported within 2 scan intervals following last adequate radiological tumor assessment	Date of death	Event
Tumor progression (per RECIST 1.1) documented within 2 scan intervals following previous adequate radiological tumor assessment	Earliest of the target, non-target and new tumor assessment dates	Event
Tumor progression (per RECIST 1.1) documented after 2 scan intervals following previous adequate radiological tumor assessment	Date of previous adequate radiological assessment	Censored
New anticancer treatment started and no tumor progression	Date of previous adequate radiological assessment immediately prior to start of new therapy	Censored
No tumor progression (per RECIST 1.1) and subject lost to follow-up or withdrawal of consent	Date of last adequate radiological Assessment	Censored

Notes: (1) Symptomatic deteriorations (i.e. symptomatic progressions, which are not radiographically confirmed) will not be considered as progressions.
(2) If target, non-target and new lesion assessments have different dates within a visit, then the earliest of those dates will be considered as the date of the tumor assessment if the assessment for that visit is progressive disease (PD); otherwise the latest date will be used.
(3) Adequate radiographical tumor assessment refers to an assessment with overall response of CR, PR, SD or PD.

4.9.1.3.3 Best Overall Response

All subjects will have their response classified using their Best Overall Response (BOR). Classification will follow RECIST (Version 1.1) criteria. In order to classify a BOR as stable disease (SD), the assessment must be made a minimum of 6 weeks from baseline. Otherwise, the BOR will be not evaluable (NE), unless any PD was further documented, in which case BOR will be classified as PD.

The objective response rate (ORR) is defined as the number of subjects who had a BOR of CR, or PR divided by the number of subjects in the population under study.

Disease control rate (DCR) is defined as the number of subjects who had a BOR of CR, PR or SD divided by the number of subjects in the population under study.

4.9.1.3.4 Quality of Life

The Quality of Life (QoL) assessment will be conducted using the EORTC-QLQ-C30 (version3), a self-administered, cancer-specific questionnaire with multidimensional scales. The 30 items (Q1-Q30) of the QLQ-C30 are scored to obtain 15 scales (one global health status/QoL scale, five functional scales, and nine symptom scales/items). A linear transformation will be used to obtain scales ranging from 0 to 100.

The scoring for the EORTC QLQ-C30 is detailed in QLQ-C30 Scoring Manual. For multiple-item scales, missing items will be imputed based on the mean of the completed items if ≥50% of contributing items are completed. No other adjustment or imputation for missing data will be performed. Refer to Section 4.11.3.3 for details of analysis.

4.9.1.4 Multiple Comparisons/Multiplicity

Inferential assessment of secondary outcomes will take place only if the primary outcome (OS) is found to differ significantly between treatments when analyzed via the primary analysis method. The secondary outcomes will be inferentially assessed using a hierarchical analysis method (e.g. inferential analysis will stop and not continue to subsequent secondary outcomes after obtaining a probability greater than or equal to two-sided 0.05). The order by which the secondary outcomes will be tested is outlined in section 4.9.3 Sensitivity analyses will be done on an exploratory basis and are not included in the hierarchical testing for secondary outcomes.

The study endpoints (OS/PFS/ORR/DCR) based on pre-defined biomarker population will be removed from the hierarchical test order of the secondary endpoints.

The hierarchical testing flow can be understood as follows:

If the p-value of the primary outcome analysis is less than required based on interim analysis conducted during the trial or the final analysis at the trial, then superiority of BBI608 in terms of OS in ITT population can be concluded. Next, the first secondary outcome of the hierarchy will be evaluated for statistical significance.

If the p-value of the first secondary outcome analysis is also found to be below the adjusted two-sided alpha of 0.05, then superiority of BBI608 in terms of OS in the biomarker positive population can be concluded. Next, the second secondary outcome of the hierarchy will be evaluated for statistical significance. This process will continue for all secondary outcomes in the hierarchy, unless an analysis of a secondary outcome reveals a p-value greater than or equal to the adjusted two-sided alpha of 0.05.

In such an event, the hypothesis associated with the secondary outcome will be concluded to be non-statistically significant, and inferential testing for the subsequent secondary outcomes in the hierarchy will be terminated. Hence, all p-values found in the analyses of the remaining secondary outcomes in the hierarchy will be considered nominal.

As QOL is an exploratory analysis, alpha and all p-values will be nominal, and conclusions related to statistical significance cannot be established.

Refer to section 4.9.1.5 for the adjustment of significance level due to interim analysis. All analyses described in this section are based on availability of data.

4.9.1.5 Interim Analysis

There will be one interim analysis to stop this study early for success presented to the independent Data Safety and Monitoring Board (DSMB). The interim analysis will be performed on OS, the primary endpoint of this study, when at least 2/3^{rds} of the required number of events (380) have been observed in the ITT population.

The cut-date for the interim OS analysis will be the date of the 380th death. Overall Survival will be derived based on the data up to cutoff.

If the interim analysis does not provide evidence of superiority, the study will continue until final analysis at 566 events. This analysis will be based on the stratified log-rank test adjusting for the stratification variables at randomization, and will test the following:

H_0: survival on BBI608 + paclitaxel = survival on placebo + paclitaxel

vs.

H_1: survival on BBI608+paclitaxel > survival on placebo + paclitaxel

The comparison of the overall population will be tested based on a generalization of the Lan-DeMets error spending function approach using an O'Brien-Fleming stopping boundary to reject H_0, controlling for a two-sided alpha of 0.05 of overall study. For example, if exactly 380 deaths (2/3^{rds}of the required events) were in the locked database for the interim analysis, the nominal critical point for rejecting H_0 would be 2.502, corresponding to a two-sided p-value of 0.012. Superiority of BBI608 may be declared, early, if the two-sided p-value from stratified log-rank test is ≤0.012 with survival being greater in the BBI608 plus weekly paclitaxel treatment group compared to the placebo plus weekly paclitaxel treatment group. Sample size calculations were conducted using East version 6.2.

The nominal critical point at the final analysis for rejecting H_0 would be 1.994 corresponding to a two-sided p-value of 0.0462.

Information rate	bounds reject H_0	sign.level one-sided	sign.level two-sided	events
0.67	2.502	0.006	0.0124	380
1.0	1.994	0.0231	0.0462	566

Should the trial not stop for efficacy at this interim analysis, the final analysis will occur when 566 deaths are observed. The nominal critical value for rejecting H0 at this final analysis would be 1.994, corresponding to two-sided p-value of 0.0462.

Correlative analysis and endpoints of biomarker positive population won't be included at the interim analysis. At interim analysis, the endpoints (OS/PFS/ORR/DCR) based on pre-defined biomarker population will be removed from the hierarchical test order of the secondary endpoint.

As detailed in section 4.10.9, the DSMB will meet every 6 months, primarily to evaluate safety of the trial. For each DSMB meeting, Open Reports, as detailed below, will be provided by the study biostatistician. The reports should be provided to DSMB members not later than 5 working days prior to the date of the meeting.

Open Reports, available to all participants attending the DSMB meeting, can include information such as data on recruitment, demographics, baseline characteristics, protocol compliance, site performance, monitoring issues and resolutions thereof, and general (ungrouped) adverse events.

Closed Reports, available only to participants attending the Closed Sessions of the DSMB meeting, will include analyses in line with the needs of the DSMB.

The following data will be presented in Open and Closed Sessions:

- Subject disposition (including dates of informed consent, randomized, dosed, withdrawal/completion, primary reason for withdrawal, etc)
- Demographics
- Extent of exposure of study drug
- AEs, including treatment-emergent AE, SAEs
- Medical history
- Concomitant medication
- Protocol Deviations

The DSMB will receive blinded data listings sorted by treatment groups (e.g. labelled A & B) that do not disclose the treatment arm. If the DSMB decides that it is necessary to analyse unblinded data, such unblinded data will be made available. Detailed subject listings will contain all relevant information to monitor the safety of the study subjects. After agreement on the format, the final listing details will be added as an attachment to the DSMB Charter (see SAP Section 3.1).

The DSMB members will receive both blinded and unblinded summary tables regarding: demography, any protocol deviations, SAEs, AEs, deaths, patients who withdraw from study treatment and/or are lost to follow up. Interim safety analyses will be performed using the safety population.

4.9.1.6 Examination of Subgroups

A subset analysis will be performed as a sensitivity analysis for the primary endpoint, overall survival (OS), to address the benefit of BBI608 between the groups across the following factors:

- Prior taxane therapy (yes; no)
- Disease measurability by RECIST 1.1 (measurable disease present; not present)
- Time to progression on first line therapy (<6 months; ≥6 months from start of first line therapy)

- Geographical region (Asia; North America, Europe and Australia; South America)
- Race (white; black; Asian; other)
- Number of previous chemotherapy drug classes [platinum, fluropyrimidine, taxane, anthracyclines], including during adjuvant or neoadjuvant therapy (<3 *versus* ≥3)
- Age (<65 *versus*≥65)
- Sex (male *versus* female)
- Prior trastuzumab-containing therapy (yes *versus* no)
- Type of cancer (gastric *versus* GEJ)
- Prior radiotherapy (yes *versus* no)
- Presence of peritoneal metastases (yes *versus* no)
- Histologic type (diffuse, intestinal, mixed, other)
- Received post protocol anticancer treatment [i.e. 3rd line regimen] (yes *versus* no)
- Number of organ sites involved at baseline (≥2 *versus*<2)
- Presence of primary tumor (present *versus* absent)

Further details on the subset analysis can be found in section 4.9.2.2.

An additional subset analysis, as described above, will be performed for all secondary efficacy endpoints that are found to be statistically significant following hierarchical testing.

In addition, the relationship between the primary endpoint (Overall Survival) and binary biomarkers will be also examined using the log-rank test, stratified by treatment, and baseline ECOG performance status (0 vs 1), geographical region (Asia vs North America, Europe and Australia vs South America), time to progression on first line therapy (< 6 months vs. > 6 months), disease measurability by RECIST 1.1 (measurable disease present vs. not present and prior taxane therapy (yes vs. no). Subgroup analysis by the status of the binary biomarker may also be performed.

The relationships between the binary response variable (e.g. objective response, DCR) with baseline values of biomarkers as well as with their change from baseline at different time points will be investigated using logistic regression that includes assigned treatment, biomarker value and treatment-by-biomarker interaction, stratified by baseline ECOG performance status (0 vs 1), geographical region (Asia vs North America, Europe and Australia vs South America), time to progression on first line therapy (< 6 months vs. > 6 months), disease measurability by RECIST 1.1 (measurable disease present vs. not present and prior taxane therapy (yes vs. no). An additional model that includes prognostic factors may be investigated. The relationship between response and binary biomarkers (e.g. baseline phospho-STAT3 status or STAT3 status) will also be examined using the Cochran-Mantel-Haenszel (CMH) test, stratified by treatment, baseline ECOG performance status (0 vs 1), geographical region (Asia vs North America, Europe and Australia vs South America), time to progression on first line therapy (< 6 months vs. > 6 months), disease measurability by RECIST 1.1 (measurable disease present vs. not present and prior taxane therapy (yes vs. no). Subgroup analysis by the status of the binary biomarker may also be performed.

4.9.2 Primary Efficacy Variable

4.9.2.1 Main Analysis

The primary analysis for the assessment of efficacy will evaluate Overall Survival in the Intent-to- Treat (ITT) study population. All randomized subjects will be included in the primary analysis of OS, which is defined as the time interval between the date of randomization and the date of death from any cause. If a subject is not known to have died on or before the date of data cut-off, OS data will be censored on the last date (on or before the cut-off date) the subject was known to be alive. Subjects who lack data beyond randomization will have their survival times censored on the date of randomization. Subjects will be analyzed in the arm to which they are randomized regardless of the treatment they received.

The survival experience of subjects in both treatment arms will be summarized by the Kaplan-Meier method and compared primarily by a stratified log-rank test adjusting for the randomization stratification variables listed below. If any randomization stratum proves to have an inadequate sample size, it will be dropped from the primary analysis or pooling will be applied. All of the stratification variables will be included into a single model and will be categorized as follows:

- Geographical region (Asia *vs* North America, Europe, and Australia *vs* South America)
- Time to progression on first line therapy (<6 months *vs* ≥6 months from start of first line therapy)
- Disease measurability by RECIST 1.1 (measurable disease present *vs* not present)
- Prior taxane therapy (yes *vs* no)

The resulting test statistic and p-value for the stratified log rank test will be reported. In addition, Kaplan-Meier curves will be generated for OS and used to calculate quartiles and 95% CIs of the median OS for each treatment arm in the Intent-to-Treat Set. In addition, Kaplan-Meier estimates of the OS rate at fixed time points (e.g. 8, 16, 24, 32 weeks) along with their two-sided 95% CIs will also be derived from the Kaplan-Meier estimation. The CI for OS rate will be calculated by applying asymptotic normality to the log-log transformation of OS rate which is the default method to calculate the CI for the survival function as implemented within PROC LIFETEST in SAS software. A Kaplan-Meier plot of overall survival by treatment arm will be provided.

A listing of overall survival will be provided by treatment arm.

4.9.2.2 Sensitivity Analyses

1. The Overall Survival analysis will be repeated using the per-protocol population and safety analysis population.

2. A stratified Cox proportional hazard model will be run adjusting for the randomization stratification variables. In addition tothe treatment factor (BBI608 + paclitaxel *versus* placebo + paclitaxel), the following factors at subject entry will be included in the stratified Cox proportional hazards model:
    - Number of previous chemotherapy drug classes [platinum, fluropyrimidine, taxane, anthracyclines], including during adjuvant or neoadjuvant therapy (< 3 *versus*≥ 3)
    - Age (< 65 *versus*≥ 65)

- Sex (male *versus* female)
- Prior trastuzumab-containing therapy (yes *versus* no)
- Location of primary tumor (gastric *versus* GEJ)
- Prior radiotherapy (yes *versus* no)
- Presence of peritoneal metastases (yes *versus* no)
- Histologic type (diffuse, intestinal, mixed, other)
- Number of organ sites involved at baseline (≥2 *versus*<2)
- Presence of primary tumor (present *versus* absent)

The hazard ratio (HR) and its 95% CIs will be estimated for all factors.

3. Exploratory models: Two exploratory models will be generated. The first will evaluate whether study outcomes could have been influenced by receipt of post-protocol anti-cancer treatment. The second will evaluate the effect of treatment versus placebo in those who receive minimum effective treatment, defined in Section 4.5.

4. The subset analysis outlined in section 4.9.1.6 will performed on the primary endpoint (OS) utilizing Cox proportional hazards models. Each subset analysis will contain treatment group and the variable of interest as factors. No adjustments for covariates will be made. For each subset analysis, the HR and 95% CI's will be estimated for OS between the treatment groups across the different subsets. A forest plot of overall survival will be presented by subgroup.

No control for type I error will be implemented for the sensitivity and subset analyses outlined above.

4.9.3 Secondary Efficacy Variables

The following secondary outcomes will be inferentially assessed using a hierarchical analysis method (e.g., inferential analysis will stop and not continue to subsequent secondary outcomes after obtaining a p-value greater than 0.05). The order of endpoint presentation below reflects the hierarchy of inferential testing. Inferential assessment of secondary outcomes will take place only if the primary outcome (OS) is found to differ significantly between treatments when analyzed via the primary analysis method detailed above.

1. *Overall Survival in the Predefined Biomarker-positive Population*

   Overall Survival (OS) in biomarker-positive population is defined as the time from randomization to death due to any cause. The analyses and displays for OS in the ITT population, as described in section 4.9.2.1, will also be performed for OS in the biomarker-positive population, using similar methodology.

2. *Progression-Free Survival in the ITT General Study Population*

   Progression-Free Survival (PFS) in the ITT population is defined as the time from randomization to the first objective documentation of disease progression or death due to any cause. If a subject

has not progressed or died at the time of final analysis, PFS will be censored on the date of the last tumor assessment. This includes subjects who have withdrawn their consent from study procedures and/or follow-up. The analyses and displays for OS in the ITT population, as described in section 4.9.2.1, will also be performed for PFS in the ITT population, using similar methodology.

3. *Progression-Free Survival in the Predefined Biomarker-positive Population*

   PFS in biomarker-positive population is defined as the time from randomization to the first objective documentation of disease progression or death due to any cause. The analyses and displays for OS in the ITT population, as described in section 4.9.2.1, will also be performed for PFS in the biomarker-positive population, using similar methodology.

4. *Objective Response Rate in the ITT General Study Population*

   Objective Response Rate (ORR) is defined as the proportion of subjects with a documented complete response or partial response (CR + PR) based on RECIST 1.1. The primary estimate of ORR will be based on subjects with measurable disease by RECIST 1.1 at randomization. CMH test will be used for comparison of the two treatment groups, adjusting for the randomization stratification variables (outlined in section 4.9.1.2). The test statistic and p-value will be reported.

In addition, a logistic regression analysis may be performed as a sensitivity analysis, with Objective Response as a binary endpoint. The effect of treatment arm will be evaluated in the model, adjusting for randomization stratification variables. Model based estimates of the proportions of subjects with objective response, as well as the corresponding 95% CI, will be presented by treatment arm. The odds ratio between the treatment arms will be estimated, and the corresponding 95% CI and p-value will be reported.

5. *Disease Control Rate in the ITT General Study Population*

   Disease Control Rate (DCR) is defined as the proportion of subjects with a documented complete response, partial response, or stable disease (CR + PR + SD) based on RECIST 1.1. The primary estimate of DCR will be based on subjects with measurable disease by RECIST 1.1 at randomization. CMH test will be used for comparison of the two treatment groups, adjusting for the randomization stratification variables (outlined in section 4.9.1.2). The test statistic and p-value will be reported.

   In addition, a logistic regression analysis may be performed as a sensitivity analysis, with Disease Control as a binary endpoint. The effect of treatment arm will be evaluated in the model, adjusting for randomization stratification variables. Model based estimates of the proportions of subject with disease control, as well as the corresponding 95% CI, will be presented by treatment arm. The odds ratio between the treatment arms will be estimated, and the corresponding 95% CI and p-value will be reported.

4.10 Safety Evaluation

All safety summaries, analyses and listings will be based upon the Safety Analysis Set as defined in Section 4.5.

In general, all safety data will be listed, sorted by treatment group, site, and subject identification number. Descriptive statistics will be calculated for quantitative safety data and frequency counts and percentages will be compiled for classification of qualitative safety data. All percentages will be calculated based on the number of subjects in the Safety Analysis Set, unless otherwise indicated. Repeated or unscheduled tests are not summarized for each scheduled visit, but are included for summaries of maximum and minimum post-baseline values, shift-table analyses, and listings. No inferential statistical analysis is planned for safety data, unless otherwise specified.

The incidence of adverse events will be summarized by type of adverse event and severity using the NCI Common Terminology Criteria for Adverse Events Version 4.0. A Fisher's exact test will be used to compare adverse events between the two arms if required.

4.10.1 Safety Data Handling and Definitions

4.10.1.1 Treatment-Emergent Adverse Events (TEAEs)

An AE will be regarded as **treatment-emergent**, if

- its onset date occurs any time on or after the date of administration of the first dose of study treatment (either BBI608 or placebo or paclitaxel) up to 30 days after the last dose of study treatment (or up to any time if serious and considered related to study treatment); or
- it occurs prior to first dose date and worsens while on therapy or up to 30 days after the last dose of study treatment (or up to any time if serious and considered related to study treatment).

4.10.1.2 Adverse Events of Special Interest (AESIs)

The list of AESIs (see table directly below) consists of pre-specified selected adverse events that are given special consideration because they have been associated with other agents in a similar class of drugs or that were observed in preclinical evaluation or earlier clinical studies of BBI608. For each AESI, a list of clinically relevant MedDRA preferred terms will pre-specified based on review of blinded data as well as review of standard MedDRA queries (SMQs).

In order to review AESIs in the most medically meaningful way selected preferred terms that represent the same phenomenon will be grouped together. The MedDRA preferred terms that are grouped under each of the AESI terms will be provided in the compound-level safety document.

System	Event
Gastrointestinal	Diarrhea Nausea(Procedural nausea, Prophylaxis of nausea and vomiting)

	Vomiting(Infantile vomiting, Post-tussive vomiting, Prophylaxis of nausea and vomiting, Vomiting in pregnancy, Vomiting projectile) Abdominal Pain(Abdominal pain lower, Abdominal pain upper) Flatulence Fecal Incontinence Blood in stool Mucositis Oral Burping Bloating
Constitutional	Fatigue(Muscle Fatigue, Respiratory Fatigue) Chills Weight Loss(Abnormal Weight Loss, Weight Decreased)
Metabolism and Nutrition	Anorexia Hypokalemia Hypomagnesemia Dehydration
Neuro-Psychiatric	Dysgusia Confusion(Confusion postoperative, confusionalstate) Dizziness
Renal and Urinary	Urine Discoloration
Skin And Subcutaneous Tissue	Rash Maculo-Papular

4.10.1.3 Consolidated Adverse Events

Consolidated terms comprising clinically synonymous MedDRA preferred terms have been defined in order to assist in identifying relevant safety differences between treatment arms. Summaries of the incidence of these consolidated terms will supplement the summaries by MedDRA preferred terms specified above.

4.10.1.4 Previous/Concomitant Medications

Medications will be coded using the WHO Drug coding dictionary version Dec2015. All of the displays will be summarized by ATC codes Level 3 and preferred term unless otherwise specified.

Medications starting after the completion/withdrawal date will be listed but will not be classified or summarized.

Medications that start and stop prior to the date of first treatment administration will be classified as ‘prior’ medications. If a medication starts on or after the date of first treatment administration, then the

medication will be classified as 'concomitant'. If a medication starts before the date of first treatment administration and stops on or after the date of first treatment administration, then the medication will be categorized as both a 'prior' and 'concomitant' medication.

Handling of partial dates will follow the conventions in Section 4.10.1.6.

If a subject takes the same medication (i.e. same preferred term) more than once, they are only counted once under the count for preferred term per period. If a subject takes more than one medication in a particular ATC Level 3 term, they will only be included once in the count for the term per period, but will appear in the count for each appropriate preferred term within the ATC Level 3 term (unless it is the same preferred term).

The incidence of concomitant medications use will be summarized by ATC Level 3 term and preferred term for each treatment group and period. The incidence will be ordered by decreasing incidence based on the BBI608 column. In addition to the incidence, this will also include a count of the total number of medications for each WHO Drug preferred term for each treatment group and period.

Incidence of most common concomitant medications will be summarized. 'Most common' will be defined as those medications occurring in ≥ 2% of subjects in the BBI608 treatment group.

All details recorded in relation to prior and concomitant medication will be listed by subject.

4.10.1.5 Partial Dates for Adverse Events and Previous/Concomitant Medications

For the subject data listings, no imputation of incomplete dates will be applied. The listings will present the incomplete dates without any change.

Dates missing the day or both the day and month of the year will adhere to the following conventions:

- The missing day of onset of an AE or start date of a therapy will be set to:
  - first day of the month that the event occurred, if the onset yyyy-mm is after the yyyy-mm of first study treatment
  - the day of the first study treatment, if the onset yyyy-mm is the same as yyyy-mm of the first study treatment
  - the date of informed consent, if the onset yyyy-mm is before the yyyy-mm of the first treatment
- The missing day of resolution of an AE or end date of a therapy will be set to:
  - the last day of the month of the occurrence. If the subject died in the same month, then set the imputed date as the death date
- If the onset date of an AE or start date of a therapy is missing both the day and month, the onset date will be set to:
  - January 1 of the year of onset, if the onset year is after the year of the first study treatment

  - the date of the first treatment, if the onset year is the same as the year of the first study treatment
  - the date of informed consent, if the onset year is before the year of the first treatment

- If the resolution date of an AE or end date of a therapy is missing both the day and month, the date will be set to:
  - December 31 of the year of occurrence. If the subject died in the same year, then set the imputed date as the death date
- If date is completely missing, then no imputation will be done and the event will be considered as treatment emergent (for AEs) or concomitant (for medications) unless the end date rules out the possibility.

4.10.1.6 Study Drug Exposure

Exposure analyses will be based on the actual dose administered (in mg) for BBI608 or body surface area (BSA in m^2) for paclitaxel. The BSA to be used for calculating each dose of paclitaxel will be calculated using the Dubois & Dubois formula (see below) based on the last available weight and height prior to each infusion.

$BSA\ [m^2] = (Weight\ [kg]^{0.425} * Height\ [cm]^{0.725}) * 0.007184$

Calculated doses will be rounded to the nearest integer.

For subjects who did not receive any amount of a study drug, the dose exposure parameters for that treatment (number of infusions, duration of treatment, cumulative dose, dose intensity, relative dose intensity) will be set to 0. Cumulative dose, dose intensity and relative dose intensity will remain missing if they cannot be derived due to missing weight, height, or BSA.

BBI608 or Placebo Treatment:

- Duration of treatment (weeks; 1 day added to duration of treatment because administration of each cycle is for 1 day) = [(Date of last dose - date of first dose) +1] ÷ 7
- Cumulative dose, dose intensity, relative dose intensity:
  - Cumulative dose (mg) = Sum of all (total dose administered [mg])
  - Weekly dose intensity (mg/week) = (Cumulative dose) ÷ (Duration of treatment)
  - Planned weekly dose intensity (mg/week) = 2 * 480 mg * 7 = 6720 mg/week
  - Relative dose intensity (%) = (Weekly dose intensity) ÷ (Planned weekly dose Intensity) *100
- Number of dose level reductions: Dose level reductions as reported in the eCRF will be calculated on a by-subject-basis and number of events of dose reduction over time.

- Dose delays: As reported in the eCRF
- Dose Omissions (Not Administered): As reported in the eCRF

Paclitaxel Treatment:

- Duration of treatment (weeks; 14 days added to duration of treatment because last administration of each cycle is for 2 weeks [on day 1, 8, 15 of each 4-week cycle]) =[(Date of last dose - date of first dose) + 14] ÷ 7
- Cumulative dose, dose intensity, relative dose intensity:
  - Cumulative dose (mg/m^2) = Sum of all (total dose administered [mg] ÷ BSA using last available weight [m^2])
  - Weekly dose intensity (mg/ m^2/week) = (Cumulative dose) ÷ (Duration of treatment)
  - Planned weekly dose intensity (mg/m^2/week) = 3*80mg/ m^2 / 4 weeks = 60 mg/ m^2/week
  - Relative dose intensity (%) = (Weekly dose intensity) ÷ (Planned weekly dose intensity) * 100

- Number of dose level reductions: Sum of the number of dose level reductions as reported in the eCRF
- Dose delays: As reported in the eCRF
- Dose Omissions (Not Administered): As reported in the eCRF

4.10.2 Extent of Exposure

The following exposure variables will be presented for BBI608/Placebo, Paclitaxel, and overall:

- Number of cycles completed, descriptively and categorically (all cycles)
- Total Dose (all cycles)
- Treatment Duration (all cycles)
- Number of doses received (all cycles)
- Dose intensity (all cycles)
- Relative dose intensity (all cycles)
- Number of subjects with dose interruptions (by cycle and overall)
- Number of subjects with dose reduction (by cycle and overall)
- Number of subjects with treatment discontinuation (by cycle and overall)

The listing will provide data collected on the Study Drug (BBI608/Placebo and Paclitaxel) Administration eCRF pages for each subject.

4.10.3 Adverse Events

Treatment-emergent adverse events (TEAEs) will be summarized by Medical Dictionary for Regulatory Activities (MedDRA™) System Organ Class (SOC) and Preferred Term (PT) classified from verbatim terms, from the latest version available. If more than one AE is recorded for a subject within any SOC or PT term, the subject will only be counted once on the most severe National Cancer Institute Common Terminology Criteria for Adverse Events (NCI-CTCAE) Version 4.0 grade and the closest relationship to treatment. AEs reported with a relationship to study medication considered by the investigator to be 'possible', 'probable' or 'definite' will be considered attributable to study medication. Missing classifications concerning study medication relationship will be considered as related to study medication by the investigator.

An overview of AE's and TEAEs will be provided. The following set of summaries will be presented across treatment groups for any TEAE:

- By SOC and PT (all grade, and grade $\geq$3)
- By PT (all grade, and grade $\geq$3) by decreasing frequency on BBI608 arm
- By consolidated terms (consolidation using BBI608 project-specific standard)
- By max grade 1-5

These summaries will also be produced for TEAEs considered related to study treatment by the investigator, for 'TEAEs considered related to BBI608' by the investigator, and for 'TEAEs considered related to paclitaxel' by the investigator.

Treatment Emergent Adverse Events of Special Interest (TEAESIs) will be summarized using MedDRA Version 19.1 PT.

All collected AEs (treatment emergent or non-treatment emergent) will be presented in a listing by treatment group, sorted by SOC.

4.10.4 Deaths, Serious Adverse Events, and Other Significant Adverse Events

All deaths that occur after the administration of the first dose of any study drug and within 30 days of last dose of any study drug from the signing of the informated consent to the end of the follow up period for death will be provided in a listing. The listing will include primary cause of death, and the number of days relative to the administration of first and last dose.

The number and percentage of subjects who died during the study treatment and within 30 days after the last dose will be presented by primary cause of death and by treatment group. A summary table will be provided for TEAEs leading to death:

- By SOC and PT
- By consolidated terms

A listing will be provided for all adverse events leading to death.

Serious adverse events (SAEs) are those events that result in death, are life-threatening, require or prolong hospitalization, result in persistent or significant disability/incapacity, cause congenital anomaly/birth defect, and/or considered significant by the investigator for any other reason. The following list of summaries will be provided for TESAEs:

- By SOC and PT
- By consolidated terms

These summaries will also be produced for 'TESAEs considered related to study treatment by the investigator', for 'TESAEs considered related to BBI608 by the investigator' and for 'TESAEs considered related to paclitaxel by the investigator'.

Additionally, a summary by PT will also be produced for 'TESAEs considered related to study treatment by the investigator', for 'TESAEs considered related to BBI608 by the investigator' and for 'SAEs considered related to paclitaxel by the investigator', which resulted in hospitalization.

A listing of all SAEs will be presented by treatment group.

TEAEs leading to discontinuation, TEAEs leading to dose reduction, and TEAE's leading to dose interruption, will be summarized:

- By SOC and PT (all grade, and grade >=3)
- By consolidated terms

Listings will be provided for AEs leading to dose interruption, AE's leading to dose reduction, and AE's leading to treatment discontinuation. These listings will be presented by treatment group.

4.10.5 Clinical Laboratory Evaluation

The following laboratory variables will be determined in accordance with the schedule of procedures as specified in subject evaluation flow sheet.

Clinial Laboratory Assessments

Hematology	Leukocytes, Hemoglobin, Platelets, Neutrophils, Lymphocyte, Monocytes, Eosinophils, Basophils, and Absolute Neutrophils
Clinical Chemistry	Potassium, Creatinine, Creatinine Clearance, Phosphate, Magnesium, Albumin, Total Bilirubin , Aspartate aminotransferase (AST), Alanine aminotransferase (ALT), Alkaline phosphatase, and Lactate Dehydrogenase (LDH).
Urinalysis	Specific Gravity, Glucose, Protein, Blood
Pregnancy Test	Choriogonadotropin Beta

Laboratory values will be converted to standard (SI) units, and the numerical results will be graded using NCI-CTCAE Version 4.0. Laboratory results not corresponding to a NCI-CTCAE Version 4.0

term will not be graded. For values such as Sodium and Potassium there will be two bi-directional parameters (hyper and hypo) created, and the tests will be graded by CTCAE v4.0 in both directions.

Clinical laboratory results will be summarized by treatment group with descriptive statistics at each cycle:

- Observed test results for each lab parameter
- Change from baseline for each lab parameter

In addition, for the graded laboratory parameters, shift from baseline will be provided by cycle and for the maximum grade on the study.
Laboratory values will be classified as normal, low, or high based on normal ranges supplied by the central laboratory. Laboratory categories will be expressed in terms of the L (below LLN), N (between LLN and ULN) and H (above ULN) classifications for numerical measurements and normal, abnormal for categorical measurements. The number and percentage of subjects with abnormal values will be summarized for each laboratory parameter by cycle and treatment group. A comprehensive listing of all hematology, available urinalysis, and biochemistry data will be provided, including the test result, normal range, and change from baseline. Values outside the normal laboratory reference ranges provided by the testing lab will be flagged as high or low in the listings and also presented in a separate listing. The table below outlines CTCAE Version 4.0 grade 3 adverse event criteria for clinical laboratory parameters.

Parameter	Unit	Grade 3 Criteria
Leukocytes	(10^9/L)	HYPO: <2 - 1
Hemoglobin	(g/L)	HYPO : <80
		HYPER: Increase in >4 gm/dL above ULN or above baseline if baseline is above ULN
Platelets	(10^9/L)	HYPO: <50 – 25.0
Lymphocytes	(10^9/L)	HYPO: <0.5 – 0.2
		HYPER: > 20
Potassium	(mmol/L)	HYPO: <3.0 – 2.5
		HYPER: >6.0 – 7.0
Creatinine	(umol/L)	HYPER: >3.0 baseline; >3.0 – 6.0 * ULN

Phosphate	(mmol/L)	HYPO: <0.6 – 0.3
Magnesium	(mmol/L)	HYPO: <0.4 – 0.3
		HYPER: >1.23 – 3.30
Albumin	(g/L)	HYPO: <20
Bilirubin	(umol/L)	HYPER: >3.0 – 10.0 * ULN
Aspartate aminotransferase	(IU/L)	HYPER: >5.0 – 20.0 * ULN
Alanine aminotransferase	(IU/L)	HYPER: >5.0 – 20.0 * ULN
Alkaline phosphate	(IU/L)	HYPER: >5.0 – 20,0 * ULN

Retest results will be reported in the listings as well.

Incidence of patients with liver function test results satisfying the drug-induced liver injury (DILI) criterion defined as (> 3x upper limit of normal [ULN] for ALT/AST, >2xULN for total bilirubin and ≤ 2xULN for alkaline phosphatase at the same time-point) will be presented by Cycle.

In addition, incidence of elevated liver function test results will be presented by elevation criterion. Elevation criteria are given as follows:

- ALT (ULN - ≤ 3xULN, > 3xULN - ≤ 5xULN, > 5xULN)
- AST (ULN - ≤ 3xULN, > 3xULN - ≤ 5xULN, > 5xULN)
- Total Bilirubin (ULN - ≤ 2xULN, > 2xULN)
- Alkaline Phosphatase (ULN - ≤ 2xULN, > 2xULN)

4.10.6 Vital Signs

Summaries of markedly abnormal vital signs parameters, including blood pressure (BP) and pulse, will be presented by treatment group and cycle. For systolic and diastolic blood pressure, shift from baseline will be provided by cycle and for the worst value on the study using the following categories:

Values for vital signs for all subjects will be presented in a listing, and subjects with markedly abnormal values will be flagged.

Markedly abnormal ranges for vital signs parameters are given in the table below.

Vital Sign Parameter	Markedly Abnormal (Low)	Markedly Abnormal (High)

Systolic BP	Absolute value ≤ 90 mmHg, or a decrease from baseline ≥20 mmHg	Absolute value ≥ 180 mmHg, or an increase from baseline ≥20 mmHg
Diastolic BP	Absolute value ≤ 50 mmHg, or a decrease from baseline ≥15 mmHg	Absolute value ≥ 105 mmHg, or an increase from baseline ≥ 15 mmHg
Pulse	Absolute value ≤ 50 bpm, or a decrease from baseline ≥ 15 bpm	Absolute value ≥ 120 bpm, or an increase from baseline ≥ 15bpm

4.10.7 Electrocardiogram (ECG) Evaluations

ECG findings will be presented using frequency counts categorized as: normal, abnormal not clinically significant, abnormal clinically significant or missing at each visit, by treatment. A by subject listing will also be provided. Unscheduled visits will be included only in the listing.

4.10.8 Other Safety Evaluations

Weight at baseline will be presented by treatment group using summary statistics. Changes from baseline to on-treatment weight assessments will be presented by treatment group and time point considering the frequency of subjects with changes falling in the following categories: <-10% (loss), ≥-10% - <10%, ≥10% (gain).

Physical examination abnormalities will be summarized by body system, treatment group and time point. In addition, observed and changes in physical examination values for all subjects will be provided in a listing. Subjects with clinically significant abnormal findings will be flagged in the data listing.

The ECOG performance status (PS) results will be summarized using frequency distributions for each scheduled visit, including also the best and worst post-baseline value. A listing of observed and changes in ECOG PS for all subjects will be provided. Subjects with deterioration to ECOG PS 4will be flagged. In addition, a graphical display will be provided for time to deterioration in ECOG PS.

Reported pregnancies, as measured by serum or urine pregnancy test, will be provided in a listing, as well.

4.10.9 Data and Safety Monitoring Board (DSMB)

The DSMB is an independent expert advisory group, charged with the responsibility of evaluating safety of the trial. It is planned to have DSMB meetings to evaluate subject safety at regular intervals of every six months during the study. During these meetings, enrollment status, AEs including serious events and events of special interest, and deaths, including deaths occurring as a result of the disease under study and deaths unrelated to disease under study, will be presented. The first meeting will take place

after the first 150 subjects (combined for both arms) have completed study (ie., the first 150 primary endpoint events have occurred).

Except as otherwise noted, the identification of AEs of special interest will be based on MedDRA preferred terms, using standardized MedDRA queries and prespecified lists of preferred terms indicative of the respective clinical conditions.

There may be further ad hoc meetings (by phone or face-to-face) to evaluate the safety of the subjects, when the internal safety review of Boston Biomedical (pharmacovigilance) detects safety risks.

Within 5 days after each meeting, the DSMB will make written recommendations to Boston Biomedical to continue the study as planned, to continue the study with a minor or a major protocol amendment or to terminate the study. (refer to DSMB Conclusion Form). Recommendations for modifications, suspension or termination should be accompanied by the minimum amount of data required for the Sponsor to make a reasoned decision about the recommendation, and the rationale for such recommendations should be clearly conveyed. Both a written recommendation and oral communication should be followed with the opportunity for questions and discussion between the DSMB and Boston Biomedical / Steering Committee.

4.11 Other Analyses

4.11.1 Health Economics

A cost effectiveness analysis will be performed independently of the reporting of this study. The details will be provided in a separate analysis plan.

4.11.2 Correlative Analyses

The purpose of these studies is to explore the relationship between biomarkers and disease response.

In addition to analysis of potential biomarkers of benefit from BBI608 therapy, an analysis of markers of resistance to BBI608 will be conducted using serial plasma samples which will be drawn at baseline as well at 4weeks post initiation of therapy. This analysis may involve SNPs, CYPS, and metabolomics.

The following analysis will be performed to investigate the relationship between endpoints and biomarker levels (e.g. STAT3, phospho-STAT3, β-catenin, etc.):

For each biomarker, Cox Proportional Hazards models will be used to model the relationship between each of the following efficacy endpoints: Overall Survival, Progression-Free Survival with baseline value of the biomarker. Each model will also include assigned treatment, interaction between treatment and biomarker, and will be stratified by baseline ECOG performance status, geographical region (Asia vs North America, Europe and Australia vs South America), time to progression on first line therapy (< 6 months vs. > 6 months), disease measurability by RECIST 1.1 (measurable disease present vs. not present) and prior taxane therapy (yes vs. no).

For biomarkers with post-baseline measurements, similar analysis will be done with change from baseline of a biomarker as a covariate, for each of the time points. An additional model that includes prognostic factors may be investigated.

The relationship between the endpoint and binary biomarkers (e.g. baseline phospho-STAT3 status, β-catenin, etc.) will be also examined using the log-rank test, stratified by treatment and baseline ECOG performance status, geographical region (Asia vs North America, Europe and Australia vs South America), time to progression on first line therapy (< 6 months vs. > 6 months), disease measurability by RECIST 1.1 (measurable disease present vs. not present) and prior taxane therapy (yes vs. no). Subgroup analysis by the status of the binary biomarker may also be performed. The Kaplan-Meier method will be used to describe overall survival, progression-free survival by the status of the biomarker (e.g. phospho-STAT3-high vs. phospho-STAT3-low) and by treatment group.

The relationships between the binary response variable (e.g. objective response, DCR) with baseline values of biomarkers as well as with their change from baseline at different time points will be investigated using logistic regression that includes assigned treatment, biomarker value and treatment-by-biomarker interaction, stratified by baseline ECOG performance status geographical region (Asia vs North America, Europe and Australia vs South America), time to progression on first line therapy (< 6 months vs. > 6 months), disease measurability by RECIST 1.1 (measurable disease present vs. not present) and prior taxane therapy (yes vs. no). An additional model that includes prognostic factors may be investigated. The relationship between response and binary biomarkers (e.g. baseline phospho-STAT3 status or STAT3 status) will also be examined using the CMH test, stratified by treatment and baseline ECOG performance status), geographical region (Asia vs North America, Europe and Australia vs South America), time to progression on first line therapy (< 6 months vs. > 6 months), disease measurability by RECIST 1.1 (measurable disease present vs. not present) and prior taxane therapy (yes vs. no). Subgroup analysis by the status of the binary biomarker may also be performed. As a sensitivity analysis, logistic regression analysis will be performed for each binary response variable, using the binary biomarkers as factors, and stratifying by treatment and baseline ECOG performance status.

Exploratory analyses, in addition to those described in this section, such as alternative modeling approaches and analyses of other biomarkers are expected and may be performed. All analyses described in this section are based on availability of data.

4.11.3 Exploratory Analyses

4.11.3.1 Response Rates in biomarker-positive population

A Cochran Mantel Haenszel (CMH) test adjusted for the stratification factors will be used to compare the differences in the ORR and DCR between the treatment groups. For each analysis, the test statistic and p-value will be reported.

- *Objective Response Rate in the Predefined Biomarker-positive Population*
  Objective Response (ORR) in the biomarker-positive population [those subjects with pSTAT3 positivity on immunohistochemical (IHC) staining of Formalin Fixed Paraffin Embedded (FFPE) archival tissue] is defined as the proportion of subjects with a documented complete response or

partial response (CR + PR) based on RECIST 1.1. The primary estimate of ORR will be based on subjects with measurable disease by RECIST 1.1 at randomization.

- *Disease Control Rate in the Predefined Biomarker-positive Population*
  Disease Control Rate (DCR) in the biomarker-positive population [those subjects with pSTAT3 positivity on immunohistochemical (IHC) staining of Formalin Fixed Paraffin Embedded (FFPE) archival tissue] is defined as the proportion of subjects with a documented complete response, partial response, or stable disease (CR + PR + SD) based on RECIST 1.1. The primary estimate of DCR will be based on subjects with measurable disease by RECIST 1.1 at randomization.

4.11.3.2 Sparse Pharmacokinetic Analysis

Exploratory analyses will be performed on the bio-analytic data obtained from sparse plasma sampling in order to characterize the population pharmacokinetics of BBI608.Demographic and pathophysiologic factors that affect plasma concentration of BBI608 and paclitaxel in this population of subjects with pre-treated advanced gastric/GEJ adenocarcinoma will be examined. The exposure-response relationship between clinical and safety endpoints and BBI608/paclitaxel exposure will also be examined.

PK analyses of Plasma samples will be performed using population pharmacokinetics obtained through sparse PK sample collection. The exposure/response relationships will be summarized. Further exploratory analyses may be performed as appropriate.

4.11.3.3 Quality of Life Analysis

SCManual QLQ-C30 will be used as a reference for scoring.

The Quality of Life (QoL) of subjects will be assessed using EORTC QLQ-30 while the subject remains on BBI608/Placebo. The EORTC QLQ-30 is a self-administered cancer specific questionnaire with multi-dimensional scales. It consists of both multi-item scales and single item measures, including five functional domains, a global quality of life domain, three symptom domains, and six single items. Scoring of the EORTC QLQ-30 data will be completed following the procedures recommended by the EORTC Study Group on Quality of Life. For each domain or single item measure a linear transformation will be applied to standardize the raw score to range between 0 and 100. The quality of life data will be analyzed to look for statistically and clinically significant differences between the BBI608 *versus* placebo groups. Mean baseline scores for each subscale and summary scores will be calculated.

Compliance rates for each PRO instrument will be calculated at each assessment time point. Compliance at an assessment time point is defined as the number of subjects who were assessed for that PRO instrument divided by the expected number of subjects at that time point. The expected number of subjects:

- at baseline is equal to the number of subjects randomized
- at any post-baseline visit is equal to the number of subjects who are alive and have not progressed

A subject who answers at least one item at a time point is considered to have been assessed. Reasons for non-compliance will be tabulated by treatment arm and by assessment time point.

The endpoints in QoL analysis are the mean EORTC QLQ-C30 QoL change scores from baseline at cycle 2 (~8 weeks) and cycle 4 (~16 weeks) for the physical function and global health status/quality of life subscale scores. For both subscales, the observed values and changes from baseline will be summarized by treatment arm.

Wilcoxon rank sum tests will be used to compare the difference at each of these two time points between two treatment arms for each of these two subscales. For each comparison, the null hypothesis states that the observations in both treatment arms come from the same population. The alternative hypothesis states that the observations in both treatment arms do not come from the same population.

To perform the Wilcoxon rank sum test, the absolute values of a subscale's differences between baseline and the time point of interest will be ranked across both treatment arms. These rankings are then summed within each treatment group, and the lowest sum of the two is compared to a critical value. If the sum of the rankings is lower than the critical value, then the null hypothesis is rejected, and it is concluded that QoL change from baseline score is different between treatment groups (for the given time point and subscale of interest.)

The Hochberg method will be used to adjust for multiple comparisons in the primary analyses.

In addition, the proportion of subjects in either arm with at least a minimum of 10 unit(s) deterioration in change scores at both 8 and 16 weeks will be compared via Fisher's exact test. For each subscale and time point, the estimated proportions and nominal p-value will be reported.

By-subject listings of these data will be provided.

Graphical displays will be provided for change from baseline in EORTC QLQ-C30 Scores, as well as summary of EORTC QLQ-C30 response analysis. In addition, Kaplan-Meier plots will be provided to compare time to deterioration in EORTC QLQ-C30.

4.12 Determination of Sample Size

The primary study endpoint is Overall Survival (OS). The study is designed to have a power of 90% and a two-sided alpha of 5% to detect a 24% reduction in the continuous risk of death (HR 0.76, which corresponds to an increase of median survival from 7.36 to 9.67 months) in the Intention to Treat (ITT) general study population. It is estimated that 566 events will be required to detect this reduction which would be observed by randomizing 700 patients over 24 months and following them for an additional 12 months (including up to 5% of yearly dropouts).

There will be one interim analysis in this study to stop trial early for success to be presented to the independent Data Safety and Monitoring Board (DSMB). The interim analysis will be performed on OS, the primary endpoint of this study, when at least 2/3rds of the required number of events (380) has been observed.

The comparison of the overall population will be tested based on a generalization of the Lan-DeMets error spending function approach using an O'Brien-Fleming stopping boundary to reject H_0, controlling for a two-sided alpha of 5% at the end of the study. For example, if exactly 380 deaths (67% of the required events) were in the locked database for the interim analysis, the nominal critical point for rejecting H_0 would be 2.502, corresponding to a two-sided p-value of 0.012. Thus H_0 would be rejected, and superiority of BBI608 declared early, if the two-sided p-value from stratified log-rank test is $\leq$ 0.012.

The nominal critical point at the final analysis for rejecting H_0 would be 1.994 corresponding to a two-sided p-value of 0.0462.

Information rate	bounds reject H_0	sign.level one-sided	sign.level two-sided	events
0.67	2.502	0.006	0.0124	380
1.0	1.994	0.0231	0.0462	566

Should the trial not stop for efficacy at this interim analysis, the final analysis will occur when 566 deaths are observed. The nominal critical value for rejecting the H_0 at this final analysis would be 1.994, corresponding to a two-sided p-value of $\leq$ 0.0462.

4.13 Changes in the Conduct of the Study or Planned Analysis

Any changes made to the planned analyses in the Study Protocol should be described in this section giving justification for the changes.

5 REFERENCES

O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics 1979; 35(3);549-556.

Braunwald E, Fauci A, Hauser S, Jameson J, Kasper D, Longo D. Harrison's Principles of Internal Medicine.16th Edition. McGraw-Hill Medical Publishing Division, New York, 2005. Appendix A, Reference Values For Laboratory Tests, pages A1-A9.

Chau I, Norman AR, Ross PJ et al. Multivariate prognostic factor analysis and second-line treatment in locally advanced and metastatic oesophago-gastric cancer – pooled analysis of 1080 patients from three multicenter randomized controlled trials using individual patient data. ASCO Gastrointestinal Cancers Symposium; (2004a) abstract 5.

Collett D. Modeling Survival Data in Medical Research.2nd edition. London: Chapman & Hall; 2003.

Dolan P. Modeling valuations for EuroQol health states.Med Care. 1997;35(11):1095-1108.

EMEA, Guideline on the Evaluation of Anticancer Medical Products in Man, 2005.

Fayers P, Aaronson NK, Bjordal K, Curran D, Broenvold M, on behalf of the EORTC Quality ofLife Study Group. EORTC QLQ-C30 Scoring Manual. 3rd ed. Brussels (Belgium): EORTCQuality of Life Group, 2001.

Fleming, Thomas R, Interpretation of subgroup analyses in clinical trials, Drug InformationJournal, Vol. 29, pp.1681S-1687S, 1995

Guidance for Industry, Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics,USDHHS, FDA, May 2007

Irwin, J. O. (1949), The Standard Error of an Estimate of Expectational Life. Journal of Hygiene,47, 188-189.

Karrison, T. (1997), Use of Irwin's Restricted Mean as an Index for Comparing Survival inDifferent Treatment Groups-- Interpretation and Power Considerations, Controlled ClinicalTrials, 18, 151-167.

Meier P, Karrison T, Chappell R, Xie H. (2004). The Price of Kaplan-Meier, Journal of theAmerican Statistical Association, 99 (467):890-896

Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, PetoJ, Smith PG. (1977). Design and analysis of randomized clinical trials requiring prolongedobservation of each subject. II. Analysis and examples. Br J Cancer, 35(1):1-39.

6 APPENDICES

Appendix 1: Patient Evaluation Flow Sheet

Tests & Procedures	Pre-Treatment	During Protocol Treatment (+/-3 days)								After Protocol Treatment Discontinuation (+/- 3 days)	
		run-in[1]		Cycle 1			Additional Cycles				
Day		day 1	day 2	1	8	15	1	8	15		
Timing	≤14 days prior to randomization									4 weeks post protocol treatment	Every 8 weeks thereafter[15]
History[2] and Physical Exam	X						X			X	
ECOG PS	X						X			X	
Weight	X						X			X	
Height	X										
Vital signs	X						X			X	
Paclitaxel Infusion[3]				X	X	X	X	X	X		
Begin BBI608/Placebo administration		X									
Hematology[4,5]	X			X	X	X	X	X	X	X	
Biochemistry[5]	X						X			X	
Urinalysis[5]	X						X			X	
ECG (12-lead)	X			X						X	
Radiology and Imaging[6]	X	Every 8 weeks until progressive disease is documented									
Submission of representative block of diagnostic tumor tissue	upon request after randomization										
Blood collection for correlative studies[7]	X						X			X[8]	
Blood collection for sparse PK analysis					X	X	X				
Pregnancy test, serum or urine (if applicable)[9,10]	X						X			X	
Adverse Event assessment[11,12]	X	X	X[12]	X	X	X	X	X	X	X	
Quality of Life assessment (EORTC QLQ-C30)[13]	X						X			X[14]	
Assessment for survival of patient										X[16]	X[16]

1 BBI608/Placebo administration will begin 2 calendar days prior to the paclitaxel infusion on day 1 of cycle 1. These two days are referred to as *run-in day 1* and *run-in day 2*. *Run-in day 1* should occur within 2 working days of subject randomization.

2 Medical history must include date of diagnosis including histological documentation of malignancy, documentation of *Her2* status of tumor(if available), prior anticancer therapy and prior date(s) of disease progression.

3. Paclitaxel administration should proceed according to institutional standard practice (with respect to pre-treatment laboratory evaluation, clinical assessment, pre-medication, and monitoring during and after infusion).

4 Hematology should be done within 72 hours prior to paclitaxel administration.

5Laboratory investigations should be performed within72 hours prior to Day 1 of each Cycle of protocol treatment.

6 Tumor measurement and evaluation by RECIST 1.1 criteria. The same method of assessment and the same technique should be used to identify and

report each lesion at baseline and at reassessment during treatment. Tumor evaluations will continue until progressive disease is documented (as described in section 10). For subjects who remain on protocol therapy after objective disease progression has been documented, no further imaging assessments are mandated, but where these occur as a component of care, tumor measurements and assessment must be reported. Tumor assessments should be obtained within +/- 5 days of protocol specified schedule.

7 Sample collection should be performed at baseline and at 4weeks after randomization.

8 A sample will be collected following protocol treatment discontinuation if discontinuation occurs prior to 4 weeks of therapy.

9 In women of childbearing potential only. The minimum sensitivity of the pregnancy test must be 25 IU/L or equivalent units of HCG. Baseline pregnancy test should be done within 5 days of randomization.

10 In women of childbearing potential only a negative pregnancy test must be demonstrated every 4 weeks until 4 weeks after the administration of the final dose of protocol therapy.

11 Adverse events will be recorded and graded according to the NCI Common Terminology Criteria for Adverse Events version 4.0 (see Appendix III).

12 Adverse event assessment by phone should be performed on *run-in day 2*.

13 To be completed in clinic. Questionnaires should be completed at baseline and At 4, 8, 12, 16 and 24 weeks after randomization for as long as subject remains on Protocol therapy or until deterioration to ECOG PS 4 or hospitalization for end of life care.

14 EORTIC QLQ-C30 questionnaire will be collected in the post-protocol discontinuation period only if the subject discontinues protocol treatment prior to 24 weeks of therapy and has an ECOG PS of less than 4 and has not been hospitalized for end of life care.

15 After the first visit at which the subject has been off protocol treatment for 4 weeks, subjects will be assessed every 8 weeks for survival. Medical history at post-progression follow up must include post-protocol treatment cancer therapies.

16 In the event that the subject is unable to attend clinic, post-progression follow-up may be by means of telephone contact.

Appendix 2: Inclusion and Exclusion Criteria Resulting in the Exclusion of Patients from PP Population

This will be finalized and provided in a separate document prior to the database lock and unbinding.

Appendix 3: Prohibited Treatments Resulting in Exclusion of Patients from PP Population

Non-Permitted Treatments:
Concurrent chemotherapy, hormonal therapy (except corticosteroids), immunotherapy, biologic therapy OR other experimental agents should not be given to study patients while on protocol treatment.

Appendix 4: Scoring Details for EORTC QLQ-C30/ Windows/ Response Analysis

The scoring method for EORTC QLQ-C30 is summarized below. In this summary Qi refers to the i-th question on the QLQ-C30.

Functional scale's scores:

- Physical functioning: (1 - ((Q1+Q2+Q3+Q4+Q5)/5 -1)/3) * 100
- Role functioning: (1 - ((Q6+Q7)/2-1)/3) * 100
- Emotional functioning: (1 - ((Q21+Q22+Q23+Q24)/4-1)/3) * 100
- Cognitive functioning: (1 - ((Q20+Q25)/2-1)/3) * 100
- Social functioning: (1 - ((Q26+Q27)/2-1)/3) * 100

Global health status score:

- Global QOL: ((Q29+Q30)/2-1)/6 * 100

Symptom scale's scores:

- Fatigue: ((Q10+Q12+Q18)/3-1)/3 * 100
- Nausea and vomiting: ((Q14+Q15)/2-1)/3 * 100
- Pain: ((Q9+Q19)/2-1)/3 * 100
- Dyspnea: ((Q8-1)/3 * 100
- Insomnia: (Q11-1)/3 * 100
- Appetite loss: (Q13-1)/3 * 100
- Constipation: (Q16-1)/3 * 100
- Diarrhea: (Q17-1)/3 * 100
- Financial difficulties: (Q28-1)/3 * 100

Missing items on the scale will be handled by the following methods: values will be imputed for missing items by "assuming that the missing items have values equal to the average of those items which are present" for any scale in which at least half the items are completed. A scale in which less than half of the items are completed will be treated as missing.

The analyses of quality of life data will be restricted to intent-to-treat (ITT) patients who have a measurement at baseline and at least one measurement after baseline.

The QoL assessment is performed prior to randomization and during protocol treatment. Since exact time of assessment may vary from subject to subject, it is necessary to provide a window for each QoL time point. What follows is a description of how to assign a questionnaire to a discrete time point:

Week from randomization = (assessment date – randomization date) / 7;

Time Point	Windows
Baseline	14 days prior to or on the randomization
Week 4	0 -< 7 weeks
Week 8	7 weeks -< 10 weeks
Week 12	10 weeks -< 14 weeks
Week 16	14 weeks -< 20 weeks.
Week 24	20 weeks -< 28 weeks.

If more than one questionnaire is available for the baseline window, then the latest non-missing measurement, per question, will be considered. If more than one questionnaire is available at a time point other than baseline, then the average (per question) of the non-missing measurements will be used.

Summary statistics will be based on changes of the quality of life scores from baseline.
Baseline and Change Scores for All QoL Scales and Time Points Descriptive statistics for EORTC QOL-C30 scores (mean, standard deviation) will be presented for each scale at baseline. The same statistics will be generated at each time of post-baseline evaluation. The change scores from baseline at each time of post-baseline evaluation between treatment groups will be assessed using a Wilcoxon rank sum test for each EORTC-QLQ-C30.

QOL Response Analysis for EORTC QLQ-C30
For EORTC-QLQ-C30, QoL response for functional scales and global health status is calculated as follows: A change score of 10 points from baseline is defined as clinically relevant. Patients are considered to have clinical improvement if reporting a score 10-points or better than baseline at any time of QoL assessment. Conversely, patients are considered worsened if reporting a score minus 10-points or worse than baseline at any time of QoL assessment without any clinical improvement. Patients whose scores are between 10-point changes from baseline at every QoL assessment will be considered as stable. In contrast to functional scales, for the determination of patient's QoL response, classification of patients to improved and worsened categories is reversed for symptom scales. A Chi-square test will then be performed to compare the distributions of these three categories between two arms.

Appendix 5: Clarification for Strata Covariates Derivation and Pooling

Patients who are incorrectly stratified at the time of randomization, either because of information available subsequent to randomization or due to clerical error, the actual stratification as recorded in the CRF for these variables will be used in the primary analyses.

Randomization stratification factors from IVRS and actual stratification factors from eCRF and the discrepancy will be summarized as well.

If any randomization stratum proves to have an inadequate sample size, it will be dropped from the primary analysis or pooling will be applied. Details will be determined at blinded data review meetings prior to data base lock.

The Cox regression model, stratified for actual stratification factors will be fitted, estimated hazard ratio and 2-sided 95% CI will be provided.

Appendix 6: Clarification on Object Response Rate and Disease Control Rate: Population and Analysis Details

The objective response rate (ORR) is defined as the number of subjects who had a best overall response (BOR) of complete response (CR), or partial response (PR) divided by the number of subjects in the ITT population of patients **with measurable disease** under study. ORR will be summarized for each treatment along with the corresponding exact 2-sided 95% CI using a method based on F distribution. The treatment difference of ORR and its 95% CI based on normal distribution will be provided. CMH test will be used for comparison of the two treatment groups, adjusting for the actual stratification variables (outlined in section 4.9.1.2 except for measurable disease present or not) and p-value will be reported.

Disease control rate (DCR) is defined as the number of subjects who had a BOR of CR, PR or SD divided by the number of subjects in the ITT population of patients **with measurable disease** under study. Similar analysis as for ORR described in the above paragraph will be performed.

In addition, a stratified logistic regression analysis may be performed as a sensitivity analysis, with Objective Response as a binary endpoint. The effect of treatment arm will be evaluated in the model, adjusting for actual stratification variables (outlined in section 4.9.1.2 except for measurable present disease or not). The odds ratio between the treatment arms will be estimated, and the corresponding 95% CI and p-value will be reported.

Stratified logistic regression odds ratios will be estimated using PROC PHREG in SAS. A dummy time variable will be created, where all responders will be classified as events with an arbitrary time = t_0, and non-responders as censored with time t_1, where $t_1 > t_0$. The DISCRETE option will be used for tied observations.

Appendix 7: Clarification on Removal of Consolidated Adverse Events

Consolidated adverse events are removed and will not be analyzed in this study.