STATISTICAL ANALYSIS PLAN SUMMARY

Study Title:	**A Randomized, Single-Blinded, Phase 3 Study of Second- or Third- Line Chemotherapy with Docetaxel + Plinabulin Compared to Docetaxel + Placebo in Patients with Advanced Non-Small Cell Lung Cancer with at Least One Measurable Lung Lesion (DUBLIN-3)**
Sponsor	**BeyondSpring Pharmaceuticals, Inc.**
28 Liberty Street, 39th Floor	
New York, NY 10005	
Drug Product:	**Plinabulin (BPI-2358)**
Protocol Number:	**BPI-2358-103**
Phase:	**Phase 3**
Analysis Plan Version	**Final**
Analysis Plan Date	**Mar 20, 2021**
Author	**David Feng, MS, Senior**

LIST OF ABBREVIATIONS

AE	Adverse Event
ANC	Absolute Neutrophil Count
ANCOVA	Analysis of Covariance
ATC	Anatomical Class (from WHODRUG dictionary)
CRF	Case Report Form
CI	Confidence Interval
CMH	Cochran-Mantel-Haenszel
CR	Complete Response
CS	Clinically Significant
CSR	Clinical Study Report
CTCAE	Common Toxicity Criteria for Adverse Events
D	Docetaxel (and Placebo)
DoR	Duration of Response
DP	Docetaxel and Plinabulin
ECG	Electrocardiogram
GCP	Good Clinical Practices
HR	Hazard Ratio
ICH	International Committee for Harmonization
ITT	Intent to Treat
IV	Intravenous(ly)
K-M	Kaplan-Meier
LD	Longest Diameter
MedDRA	Medical Dictionary for Regulatory Activities
NSCLC	Non-Small Cell Lung Cancer
ORR	Objective Response Rate
OS	Overall Survival
PD	Progressive Disease
PFS	Progression-Free Survival
PK	Pharmacokinetics
PP	Per Protocol
PR	Partial Response
PT	Perferred Term
RECIST	Response Evaluation Criteria In Solid Tumors
RoW	Rest of World
RTF	Rich Text Format
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SAS	Statistical Analysis System
S.D.	Standard Deviation
SD	Stable Disease
SOC	System Organ Class
SP	Safety Population
TEAE	TreatmentEmergent Adverse Event
WHODRUG	World Health Organization Drug Reference Listing

1 INTRODUCTION

This document outlines the statistical methods to be implemented during the analyses of data collected within the scope of BeyondSpring Pharmaceuticals, Inc. Protocol BPI-2358-103 titled "*A Randomized, Single-Blinded, Phase 3 Study of Second- or Third- Line Chemotherapy with Docetaxel + Plinabulin Compared to Docetaxel + Placebo in Patients with Advanced Non-Small Cell Lung Cancer with at Least One Measurable Lung Lesion (DUBLIN-3)".*

This Statistical Analysis Plan (SAP) was prepared in accordance with the planned protocol BPI-2358-103, Amendment 11.

The purpose of this SAP is to provide a framework in which answers to the protocol objectives may be achieved in a statistically rigorous fashion, without bias or analytical deficiencies, following methods identified prior to database lock. Specifically, this plan has the following purposes:

- To prospectively outline the specific types of analyses and presentations of data that will form the basis for conclusions.
- To explain in detail how the data will be handled and analyzed, adhering to commonly accepted standards and practices of biostatistical analysis in the pharmaceutical industry. Any deviations from these guidelines must be substantiated by sound statistical reasoning and documented in writing in the final clinical study report (CSR).

2 OVERVIEW OF STUDY DESIGN

This is a Phase 3, randomized, single-blinded, active-controlled trial in patients with advanced or metastatic NSCLC that has progressed after treatment with one or two, non-docetaxel containing systemic therapy regimen(s) and/or a PD-1/PD-L1 checkpoint inhibitor therapy for advanced or metastatic disease and with at least one measurable lung lesion.

Only 2 treatment arms will be studied: the experimental arm (docetaxel + plinabulin [DP]) and the control arm (docetaxel + placebo [D]).

Dosing Regimen

On Day 1 in a 21-day cycle, all patients (in the DP and D Arms) will receive treatment with 75 mg/m^2 docetaxel administered via intravenous (IV) infusion over 1 hour (±10 minutes). On Days 1 and 8, patients randomized to DP Arm will receive plinabulin (diluted in D5W) therapy. On Days 1 and 8, patients randomized to D Arm will receive D5W. Plinabulin and D5W will be administered via IV infusion over 60 minutes (± 10 minutes) on Day 1 beginning 2 hours (± 10 minutes) from the time the docetaxel infusion begins and again on Day 8. Dexamethasone (16 mg, 8 mg twice daily, or as per institution standard; IV or oral administration) will be given the day prior to, the day of (Day 1), and the day following docetaxel infusion (Day 2). Other premedication regimens as per institutional standard clinical practice are acceptable.

In patients experiencing drug-related Grade ≥2 treatment-emergent adverse events (except alopecia, anorexia, and fatigue) according to the CTCAE (v4.03) treatment may be delayed until the adverse event has recovered to ≤Grade 1. Safety laboratory tests must meet the following criteria prior to dosing at the beginning of each subsequent cycle (AST ≤2.5 x ULN, ALT ≤2.5 x ULN (≤1.5 x ULN if alkaline phosphatase is >2.5 x ULN); serum bilirubin ≤ ULN (unless patient has Gilbert's disease, then bilirubin ≤3.0 x ULN); creatinine ≤1.5 x ULN; hemoglobin ≥9g/dL, absolute neutrophil count ≥1.5 x 10^9/L and platelets ≥100 x 10^9/L. Dose reductions may be implemented for patients who experience recurrent or specific severe toxicities. Dose reduction for docetaxel will follow the Prescribing Information for docetaxel. Dose delay and/ or dose reduction of plinabulin to 20 mg/m^2 is recommended to be strongly considered for significant tumor pain, blood pressure elevation, constipation (including ileus), nausea, vomiting, diarrhea after Cycle 1 Day 8, despite the antiemesis prophylaxis with a 5HT3 receptor antagonist, or other clinically significant AEs.

Plinabulin dosing must be withheld on a dosing day if the patient has evidence or signs of an infection on a dosing day, or any medical condition that would prevent dosing with study drug (such as an infection, Grade 4 neutropenia, or unacceptable medical condition).

Stratified randomization using blocks will be applied to randomize at a 1:1 ratio to receive docetaxel + plinabulin (DP Arm) or docetaxel + placebo (D5W) (D Arm). Stratification variables include geographic region, ECOG status; and prior lines of therapy. The details of the stratification variables will be included in Appendix A.

3 OBJECTIVES

3.1. Primary Objective

The primary objective of this study is to compare the overall survival (OS) of patients with NSCLC receiving 2nd or 3rd line systemic therapy with docetaxel + plinabulin (DP Arm) to patients treated with docetaxel + placebo (D5W) (D Arm) for advanced metastatic disease.

3.2. Key Secondary Objectives

- To compare the objective response rate (ORR) of NSCLC patients receiving 2nd- or 3rd-line systemic therapy with docetaxel + plinabulin (DP Arm) to patients treated with docetaxel + placebo (D5W) (D Arm).
- To compare progression-free survival (PFS) of NSCLC patients receiving 2nd- or 3rd-line systemic therapy with docetaxel + plinabulin (DP Arm) to patients treated with docetaxel + placebo (D5W) (D Arm).
- To compare incidence of Grade 4 neutropenia (absolute neutrophil count [ANC] < 0.5 × 109/L) on Day 8 (+/- 1 day) of Cycle 1 between the 2 treatment arms.
- To compare Month 24 OS rate of NSCLC patients receiving 2nd- or 3rd-line systemic therapy with docetaxel + plinabulin (DP Arm) to patients treated with docetaxel + placebo (D5W) (D Arm).
- To compare Month 36 OS rate of NSCLC patients receiving 2nd- or 3rd-line systemic therapy with docetaxel + plinabulin (DP Arm) to patients treated with docetaxel + placebo (D5W) (D Arm).
- To compare duration of response (DoR) of NSCLC patients receiving 2nd- or 3rd-line systemic therapy with docetaxel + plinabulin (DP Arm) to patients treated with docetaxel + placebo (D5W) (D Arm).
- To compare the mean difference in quality-adjusted time without symptoms of disease and toxicity (Q-TWiST) between patients in the DP Arm and the D Arm.
- To compare the EORTC QLQ C30 global health status/QoL and the combined symptom scales/items (excluding financial difficulties).
- To compare QoL (QLQ-LC13) in patients in the DP Arm to the D Arm
- To compare proportion of patients who received docetaxel > 8 cycles
- To compare proportion of patients who received docetaxel > 10 cycles
- To compare proportion of patients who received docetaxel > 12 cycles

3.3. Safety Objectives

- To compare the safety and adverse events profile of the DP Arm to D Arm.
- To evaluate population pharmacokinetics in patients enrolled in China and in the Rest of the World (RoW).
- To compare incidence of docetaxel dose reduction and/or docetaxel dose withheld in Cycle 2 due to Neutropenia in Cycle 1 between the 2 treatment arms.
- To compare number of docetaxel doses over all cycles
- To compare the treatment exposure-adjusted AE rates for AE of Grades 3 and 4 between the two treatment arms

- To compare the treatment exposure-adjusted AE rates for AE of Gradee 4 between the two treatment arms

4. DEFINITIONS OF PATIENT POPULATIONS TO BE ANALYZED

4.1. Intent-to-Treat (ITT) Population

The ITT population is defined as all patients who were randomized to study treatment regardless of the actual treatment received. Patients will be analyzed according to the study arm to which they were randomized.

4.2. Safety Population (SP)

The Safety Population will consist of all treated patients, including those receiving part or all of one dose classified according to the actual treatment received, regardless of random assignment.

Patients in the SP will be used for safety analysis. Patients will be analyzed according to the study medication which they have taken; that is, any patient who takes any plinabulin will be in that arm for safety analysis, even if they were not randomized to that arm.

4.3. Per-Protocol (PP) Population

The PP Population will include all patients in the ITT population, excluding those identified and documented as reportable protocol violators prior to database lock.

Patients in the PP population will also be analyzed according to the study medication which they have taken; that is, any patient who takes any plinabulin will be in that arm for PP analysis, even if they were not randomized to that arm. However, if fewer than 5% of the patients are excluded from the PP population, no analysis will be performed by PP population.

4.4. Evaluable Population for ORR

The evaluable population for analysis of the objective response rate (ORR) and best overall response, will include patients who are included in ITT, and with detectable target lesions at screening and at least one post dosing disease response assessment. Patients in the evaluable population will be analyzed according to the study medication which they have taken; that is, any patient who takes any plinabulin will be in that arm for evaluable analysis, even if they were not randomized to that arm.

5. STATISTICAL ANALYSIS CONSIDERATIONS

5.1. General Principles

Statistical analyses will be reported using summary tables, figures, and data listings. Continuous variables will be summarized with counts, means, standard deviations (S.D.), medians, minimums, and maximums. Categorical variables will be summarized by counts and by the percentage of patients. Time-to-event variables will be summarized using Kaplan-Meier (K-M) survival analysis and figures for the estimated median time.

Unless otherwise noted, data will be summarized in tabular format by treatment arms. Baseline summaries will also include a total summary column.

Formal inferential statistical analyses techniques will be discussed in subsequent sections of this SAP.

Individual patient data obtained from the case report forms (CRFs), electrocardiogram (ECG), core laboratory, pharmacokinetic (PK) data and any derived data will be presented in by-patient listings sorted by study center and patient number.

All analyses and tabulations will be performed using SAS Version 9.4 or higher on a PC platform. Table, listings, and figures will be presented in RTF format. Upon completion, all SAS programs will be validated by an independent programmer. In addition, all program output will undergo an independent senior-level statistical review. The validation process will be used to confirm that statistically valid methods have been implemented and that all data manipulations and calculations are accurate. Checks will be made to ensure accuracy, consistency with this plan, consistency within tables, and consistency between tables and corresponding data listings. Upon completion of validation and quality review procedures, all documentation will be collected and filed by the project statistician or designee.

Missing or invalid data will be generally treated as missing, not imputed, unless otherwise stated (see Section **Error! Reference source not found.**).

5.2. Population characteristics

5.2.1. Disposition patients

The following will be tabulated overall and/or by treatment arms in the ITT, SP, PP, and Evaluable populations:

- Screen failures
- Number of patients enrolled – overall and by country
- Patients disposition
- Discontinuation reasons from study treatment
- Validity of patients

Patient enrollment by site A listing will be presented to describe patient study arm, date of first and last dose, date of last visit or contact, the total number of completed cycles, and the reason for ending the study for each patient.

Listings of inclusion/exclusion criteria responses will also be provided.

5.2.2. Demographic and Baseline Characteristics

The following demographic and baseline characteristics will be summarized in the ITT, SP, PP and Evaluable population.

- Age
- Gender
- Race
- Weight

- Height
- BMI
- Body Surface Area (BSA) (≤2.4 vs. 2.4)
- ECOG status (0,1 vs. 2)
- Region (China vs. Rest of World)
- Prior PD-1/PD-L1 therapy received (Yes vs. No)
- Line of prior therapy (First Line versus Second Line)
- KRAS mutation status (Yes vs. No)
- EGFR Mutation

All demographic and baseline characteristics will be listed by study center, and patient ID.

5.2.3. Other Baseline Characteristics

The following other baseline characteristics will be summarized in ITT, SP, PP, and Evaluable population using descriptive statistics and frequency tables:

- Time since initial diagnosis
- Tumor staging
- Tumor histology type
- Tumor status (squamous vs. non-squamous)
- Extent of disease (lung. liver, adrenal gland, lymph node and other)
- Metastatic brain disease
- Prior radiotherapy
- Prior cancer therapy
- Best response from prior cancer therapy
- Prior surgical therapeutic procedure

Incomplete diagnosis dates will be imputed as detailed in Section **Error! Reference source not found.**.

Years since initial diagnosis: (Date of first dose- Date of initial diagnosis +1)/365.25.

All cancer history related data will be listed by study center, and patient number.

5.2.4. Major Protocol Violations

Major protocol violations will be identified by the clinical study team and provided to Biostatistics prior to database lock. A protocol deviation is any noncompliance with the clinical trial protocol or Good Clinical Practice (GCP). The noncompliance may be either on the part of the patient, the investigator, or the study site staff. All patients with major protocol violations will be listed by the study center and patient numbers.

5.2.5. Medical History

For Medical history, the Medical Dictionary for Regulatory Activities (MedDRA) will be used. Medical history findings (i.e., previous diagnoses, diseases or surgeries) not pertaining to the study indication, starting before the signing of the informed consent and considered relevant to the study will be tabulated by primary system organ class (SOC) and preferred term (PT) in ITT population. All the medical history will be displayed. The surgical history will be summarized in the ITT population and listed. The separate listings for previous oncology regimen and primary diagnosis will also be presented.

5.2.6. Prior and Concomitant Therapy

The following will be summarized in ITT and SP and listed:

- Prior Cancer therapy
- Prior Radiology
- Concomitant procedures
- Second line therapy

5.2.7. Concomitant medication

All medication data will be coded by drug class and indication, using the WHODrug dictionary, version 4.3. All medication taken prior to the first dose of study drug will be classified as prior medication. All medication taken on or after the first dose of study drug will be classified as concomitant medication. Medications with start and stop date that bracket the date of the first dose will be summarized as both prior and concomitant medication.

For the purpose of inclusion in the concomitant medication tables, incomplete medication start and stop dates will be imputed as detailed in Section **Error! Reference source not found.**. Based on imputed start and stop dates, medications that started on or after the date of the first dose will be included in the concomitant medications table.

Concomitant medications will be summarized in the ITT and SP by giving the number and percentage of patients by preferred term within each therapeutic class, with therapeutic class and medications in each class sorted in alphabetical order. The total number of drugs in each selected therapeutic class will also be presented, where, for example, two drugs each belonging to the same class will only contribute once to the presented count.

All prior and concomitant medications, as well as medical procedures, will be listed by the study center, and patient number.

5.2.8. Physical examination

The incidence of abnormality identified by physical examination will be summarized by the visit and by treatment arm and overall in the ITT population. Abnormal physical examination data will be displayed.

5.2.9. Study Drug Exposure

Study treatment exposure will be summarized in the SP.

The following will be summarized using descriptive statistics by study arm and overall. Also, the following will be summarized for chemotherapy administration:

- Duration of exposure calculated as (date of last dose – date of first dose+1).
- Amount of drug delivered
- Number of cycles received per patient.
- Number of cycles with dose modification and (or) dose delay.
- Reasons for dose deviations from planned therapy.

All study drug administration data will be listed by the study center and patient number.

The comparison of dose intensity of docetaxel between two arms will be made using a two-sample t-test. Dose intensity of docetaxel is the % of docetaxel administered compared to the planned dose of docetaxel per cycle. The details of the calculations of dose intensity are defined in Section **Error! Reference source not found.**.

5.3. Efficacy Analysis

5.3.1. Efficacy Endpoint

Primary Efficacy Endpoint:

- Overall survival (OS)

Secondary Efficacy Endpoints, in order of clinical relevance:

- ORR
- PFS
- Percent of patients without severe neutropenia on Day 8 of Cycle 1
- Month 24 OS rate
- Month 36 OS rate
- DoR
- DoR including SD patients
- Quality-adjusted time without symptoms of disease and toxicity (Q-TWiST)
- Quality of Life (QoL): European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30) global health status/QoL and the combined symptom scales/items (excluding financial difficulties)
- QoL: European Organization for Research and Treatment of Cancer Quality of Life Questionnaire - Lung Cancer 13 (EORTC QLQ-LC13) (Symptom combined score)
- proportion of patients who received docetaxel > 8 cycles
- Proportion of patients who received docetaxel > 10 cycles
- Proportion of patients who received docetaxel > 12 cycles
- Month 18 OS rate
- RDI for cycle 4, 6, 8, 10, 12
- Month 12 OS rate

5.3.2. Analysis of the Primary Efficacy Endpoint

The analysis of the primary efficacy endpoint will be based on the ITT population.

Overall survival is defined as the time (days) from the date of randomization to the date of death due to any cause (i.e., Date of death – date of randomization +1). In the absence of confirmation of death, survival time will be censored at the last date of follow-up when the patient was known to be alive. If the time is to be expressed in month, one month is considered 30.42 days (365/12). The hypothesis test of the primary efficacy endpoint will be performed using an overall 2-sided significance level of 0.05. The log-rank and restrictive mean survival time analyses via proc lifetest procedure in SAS with treatment as a fixed factor will be used to test the null hypothesis that OS is equal between the two treatment arms versus the alternative hypothesis that OS is different between the two treatment groups.

The relative risk reduction (RRR) will be estimated, if proportional hazard ratio is deemed a reasonable assumption, using a Cox proportional hazards (Cox PH) model with treatment as the only covariate. The point estimate and corresponding 95% confidence interval (CI) for the hazard ratio (HR, DP vs. D) will be reported. The plausibility of proportional hazards assumption will be assessed by visually comparing the plot of the log of cumulative hazard between treatments and by additionally adding a treatment by logarithm-transformed time interaction into the Cox PH model. The Efron method will be used for handling ties. The assumptions of proportional hazards will be assessed via the Standardized Score Process Test and graphical methods. If it is concluded that the assumption of proportional hazards is not justified by this data, then the data will be analyzed using Restricted Mean Survival Time method.

OS will be summarized with descriptive statistics (n, median, quartiles, number censored, 95% confidence intervals, and hazard ratios, if appropriate—proportional hazard ratio assumption holds) and mean duration of OS. Kaplan-Meier time-to-event figures will also be presented.

At Look 3 (final look), the nominal two-sided p-value will be derived. If the derived p-value is <0.04626, the null hypothesis of no treatment difference will be rejected at the two-sided significance level of 0.05 (Please refer Table 1 for more detail).

As sensitivity analyses, OS will also be analyzed for the PP and ITT populations excluding 7 mortalities which are described in Appendix B, ITT population excluding subjects who received prior PD-1/PD-L1 therapy and ITT population excluding 7 mortalities and subjects who received prior PD-1/PD-L1 therapy.

In the eventuality that a patient that has been randomized to the DP Arm, and at a later time point is taken off docetaxel due to toxicity and continues on plinabulin alone, OS will be evaluated before and after the time point that the patient continues on plinabulin alone will be reported separately in ITT as sensitivity analysis.

As an exploratory analysis, the Cox proportional hazards model, if proportional hazard ratio assumption hold, will be used for the OS analysis in the ITT population with treatment arm, ECOG performance status (0-1 vs. 2), region (China vs. RoW countries), squamous or non-squamous tumor status, prior PD-l/PD-L1 therapy received (Yes vs. No), KRAS status and line of previous therapy (First Line versus Second Line) as covariates. The Efron method will be used for handling ties. The model will also be used to estimate the hazard ratios and their two-sided 95% confidence intervals.

5.3.3. Analysis of Secondary Efficacy Endpoints

The analyses of secondary efficacy endpoints will be conducted in the ITT population. To maintain an upper boundary on the overall experiment-wise type-I error rate of 0.05, hypothesis testing of the secondary efficacy endpoints will follow a closed testing procedure. Inferential comparisons between treatment groups for one or more of the following efficacy endpoints, listed in rank order of importance, will be made provided the null hypothesis associated with the analysis of the primary efficacy endpoint (overall survival) is rejected at the two-sided 0.05 level of significance.

Inferential testing of the secondary efficacy endpoints will proceed in a sequential step-down manner, provided the null hypothesis associated with the previously tested endpoint is rejected at the two-sided significance level of 0.05. Otherwise, no further inferential testing will be conducted.

A listing of all efficacy data, including tumor measurements, will be listed sorted by study center and patient number.

5.3.3.1. Analysis of ORR

The tumor response rate is defined as the proportion of patients who have baseline measurable disease and who achieve objective complete response (CR) or partial response (PR) by RECIST criteria. Patients who do not achieve CR or PR who initiate a new cancer therapy will be counted as non-responders in the objective response analysis.

The best overall response (ORR) on the ITT population will be summarized with descriptive statistics (N, percent) and analyzed for treatment differences using Barnard's test. The Method will also be used to construct point estimates and confidence intervals for the difference 95% confidence interval of ORR will be constructed by using the Clopper-Pearson method.

5.3.3.2. Analysis of PFS

PFS is defined as the time from the date of randomization to first objective documentation of tumor progression or death due to any cause. Patients without a documented disease progression or death will be censored according to the following rules:

- If patients do not have post-baseline tumor assessment or clinically determined progression and have not died, the PFS will be censored at Day 1;
- If patients do not have documented disease progression and have not died at the time of analysis, the PFS will be censored at last adequate tumor assessment date;
- If patients are lost-to-follow-up without documented progression or death, the PFS will be censored at last adequate tumor assessment date;
- If a patient has two or more consecutive missing visits, the PFS will be censored at the time of last adequate tumor assessment no matter whether the patient has a PFS event or not after the missing visits;

Primary, sensitivity and exploratory analysis of PFS will be conducted similarly to the analysis of OS.

5.3.3.3. Analysis of Neutropenia

The treatment difference in the incidence of Grade 4 neutropenia on Day 8 of Cycle 1 will be tested using CMH Test. The Method will also be used to construct point estimates and the two-sided 95% confidence interval for the treatment difference.

5.3.3.4. Analysis of DoR

Duration of response is defined as the time from date of the first assessment demonstrating a CR or PR (whichever occurs first) to date of the first assessment demonstrating progressive disease or death. For the patients who have neither disease progression nor death before the date of database cut-off, will be censored as following:

- If patients do not have documented disease progression and have not died at the time of analysis, the DoR will be censored at last adequate tumor assessment date;
- If patients are lost-to-follow-up without documented progression or death, the DoR will be censored at last tumor adequate assessment date;
- If a patient has two or more consecutive missing visits, the DoR will be censored at the time of last complete tumor assessment no matter whether the patient has a DoR event or not after the missing visits;

Primary analysis of DoR will be conducted on the responders in ITT population.

5.3.3.5. Analysis of QoL: EORTC QLQ-C30

The EORTC QLQ-C30 will be analyzed using the summary scores. The EORTC QLQ-C30 scoring method combines the 30 questions from the questionnaire into 15 scores into three summary measures: global health status/QoL, functional scale (5 items), and symptoms scale (9 items) (EORTC 2001;Fayers et al., 2001). In addition, the EORTC QLQ-C30 Summary Score which combines which combined the symptom and functional scales (excluding the financial difficulties item)will be constructed (Hinz et al., 2012).

The global health status/QoL score, the functional score, the symptom score, and the Summary Score will be summarized at every time point of evaluation and the changes from baseline will also be summarized. For each score, the score difference between the two study arms will be tested by using a linear mixed model for repeated measures with the baseline value and treatment as covariates and no other covariates will be included in this model. The method will also be used to construct point estimates and confidence intervals. The MIXED procedure in SAS version 9.4 or later will be used for the analysis. The individual items and change from baseline will be summarized and listed by visit and treatment arms in ITT.

5.3.3.6. Analysis of QoL: EORTC QLQ-LC13

The Symptom combined score for EORTC QLQ-LC13 will be summarized at every time point of evaluation and the changes from baseline will also be summarized. Difference of symptom combined score between the two study arms will be tested by using a linear mixed model for repeated measures with the baseline value and treatment as covariates and no other

covariates will be included in this model. The method will also be used to construct point estimates and confidence intervals. The MIXED procedure in SAS version 9.4 or later will be used for the analysis. The individual items and change from baseline will be summarized and listed by visit and treatment arms in ITT.

5.3.4. Subgroup analysis

The following subgroup variables will be considered:

- Region (China vs. Rest of World)
- Prior PD-1/PDL1 therapy received (Yes vs. No)
- Line of previous therapy (First Line versus Second Line)

Analyses of primary and secondary efficacy endpoints will also be conducted in the subgroups outlined above. The results will be present descriptively (and with forest plots). The hazard ratio for the treatment effect will be estimated separately within each level of a subgroup variable using the Cox proportional hazards model with treatment as a covariate.

Additionally, homogeneity of treatment effect in subgroups, both in magnitude and direction, will be assessed by adding a covariate for the subgroup variable and the corresponding treatment-subgroup interaction to the respective Cox proportional hazards model. The Efron method will be used for handling ties.

5.4. Safety Analysis

All patients will be evaluable for safety analysis if they receive at least one dose of study drug. All patients receiving a dose of study drug will be included in all safety summaries. The safety data will be presented by study arm in individual listings and summary tables, including frequency tables for adverse events and frequency and shift tables for laboratory variables. All adverse events and abnormal laboratory variables will be assessed according to the NCI CTCAE (v 4.03) grading system. Descriptive statistics will be used to summarize ECOG performance status. Vital signs will be reported in listings and summary tables. AEs and SAEs will be reported in combined tables. However, SAEs will be tabulated in their own table as well.

All safety informations will be listed by study center and patient number.

5.4.1. Adverse Events (AEs)

For the final analyses of the safety and tolerance of study drug, all treatment-emergent and overall incidences of adverse events will be summarized by system organ class and by preferred terms (MedDRA, version 15.0 or later). All AEs will be summarized by Common Terminology Criteria of Adverse Event (CTCAE) and by MedDRA. The severity of AEs will be graded using the National Cancer Institute (NCI) v 4.03 dictionary.

Treatment-emergent Adverse Events (TEAE) are those events that occur after the first administration of any study therapy through 30 days post last dose of any active study therapy. Adverse events with missing onset dates will be summarized as TEAE regardless of severity and relationship to study medication.

An overall summary of AEs and TEAEs will be generated by treatment group and overall.

Incidences of patients with TEAEs, drug-related and/or serious TEAEs, and TEAEs causing discontinuation of study drug will be summarized by treatment grouped by MedDRA Primary System Organ Class (SOC) and Preferred Term (PT). In addition, the incidence of pre-treatment AEs and AEs during the post-study follow-up will be tabulated. The incidences of TEAEs will be presented by severity and/or NCI CTCAE adverse event grade and by relationship to study treatment.

Furthermore, all TEAEs with an incidence of greater or equal than 10% will be summarized by MedDRA for worst NCI CTCAE grade. In addition, all TEAEs with incidence of greater or equal than 5% will be summarized by MedDRA and worst CTCAE grade 3-4.

The number of patients who discontinued study treatment due to an AE or required a dose reduction or interruption caused by an AE will be summarized and patients will be listed.

Listings of AEs and laboratory toxicities by worst NCI CTCAE grade will be provided.

All information pertaining to AEs noted during the study will be listed per patient, detailing verbatim, preferred term, system organ class, start date, stop date, severity and relationship to study treatment. AE onset will be shown relative (in the number of days) to the day of the first study treatment.

5.4.2 Exposure-Adjusted Event Rates

The exposure-adjusted event rates (EAERs) for AEs of grade ≥ 3 of the DP Arm to the D Arm will be analysed using a Poisson regression model with the number of events as the dependent variable using the log of treatment exposure time as the offset variable, the treatment arm as the only covariate (i.e., without a constant). The GENMOD procedure in SAS version 9.4 or later will be used for the analysis. The Method will also be used to construct point estimates and confidence intervals for difference.

Figures for the DP Arm and the D Arm will present for each cycle, the number of patients, the number of patients with AE Grade ≥3, and the cumulative exposure in patient-years.

The exposure-adjusted event rates (EAERs) for AEs of grade = 4 of the DP Arm to the D Arm will be analysed using a Poisson regression model with the number of events as the dependent variable using the log of treatment exposure time as the offset variable, the treatment arm as the only covariate (i.e., without a constant). The GENMOD procedure in SAS version 9.4 or later will be used for the analysis. The Method will also be used to construct point estimates and confidence intervals for difference.

Figures for the DP Arm and the D Arm will present for each cycle, the number of patients, the number of patients with AE Grade = 4, and the cumulative exposure in patient-years.

According to protocol, patients in the DP arm can be given P alone after 6 cycles and with the consent of Medical Monitor. So, the patients who used P alone exposure will only be calculated based on planned P dose.

5.4.2. Deaths and SAE

The following will be produced:

- Summary table and Listing of patients who died during treatment or up to 30 days after the last dose of study treatment: patient ID, start and stop date of study treatment, date of death, cause of death

- Summary table and listing of patients who die within 60 days of the first dose of study treatment: patient ID, start and stop date of study treatment, date of death, cause of death.
- Summary table and Listing of treatment-emergent SAEs: patient ID, investigator AE term, worst grade, start and stop dates of study treatment administration, start and stop date of AE, outcome, and the action taken.
- Summary table and listing of treatment-emergent drug-related SAEs: patient ID, investigator AE term, worst grade, start and stop dates of study treatment administration, start and stop date of AE, outcome, and the action taken.

5.4.3. Clinical Laboratory Tests

Safety laboratory data will include clinical chemistries, hematology, and urinalysis. Safety summaries in the form of shift tables for key laboratory parameters showing the number and percentage of patients who experience changes in laboratory parameters during the course of the study (e.g. change from normal to high, based on the laboratory reference ranges) will be displayed. Treatment-emergent laboratory abnormalities will be summarized by the visit.

From Cycle 5 forward; central labs will not be required on the Day 8 visit; however, local labs must still be taken at least 1 day prior to infusion on Day 8. Descriptive summary statistics (mean, standard deviation, median, minimum, maximum, frequencies, and percentages, as appropriate) for laboratory values will be presented at baseline, the follow-up time points, and change from baseline for each study treatment arm

All laboratory data, values, units, normal reference range, and out-of-range flags collected in the clinical database will be included in by-patient listings for further medical review.

5.4.4. Vital Signs

Vital signs (including temperature, respiratory rate, blood pressure, heart rate, and weight) will be presented descriptively at baseline and for each follow-up time point for each study treatment arm. The number (n), mean, standard deviation, median, range will be presented. Changes from baseline to each time point will also be summarized

All vital sign parameters will be included in by-patient listings for further medical review.

5.4.5. ECGs

For patients randomized to Arm DN, the ECGs being taken just after the PK sample at approximately 10 minutes after the end of the infusion on Cycle 1 Day 1 will be analyzed for QTc prolongation by both Bazett's and Fridericia's formulae. The average of the three replicates will be used and compared to the average of the three replicates done just prior to the start of the infusion. The incidence of QTC prolongation by either calculation of >30 and >60 ms increase from baseline will be presented. The incidence of QTc prolongation by either calculation of >480 ms post-infusion will be presented.

All ECGs will be summarized descriptively using N, mean, standard deviation, median, and range for each treatment arm at each visit ECGs are collected (see Section 13).

All ECG parameters will be included in by-patient listings for further medical review.

5.4.6. ECOG

a The ECOG performance status will be summarized as shift from baseline tables by visit at all scheduled visits where performance status was assessed.

5.4.7. Neutropenia

The difference in the incidence of Grade 4 neutropenia on Day 8 of Cycle 1 will be tested using Barnard's Test. The Method will also be used to construct point estimates and confidence intervals for difference. A 95% confidence interval of proportion of Grade 4 neutropenia will be constructed by using the Clopper-Pearson method.

Additionally, the absolute neutrophil count (ANC) will be tabulated for each time point in which the data is collected. Also, the change from baseline for ANC will be summarized. The number and percentage of patients who have Grade 4 toxicity according to the CTCAE criteria for neutropenia will be summarized.

The difference of counts of neutrophil on Day 8 of Cycle 1 will be evaluated using the Wilcoxon Rank-Sum test. Hodges-Lehmann Estimation will also be used to construct point estimates and confidence intervals. The NPAR1WAY procedure in SAS version 9.4 or later will be used for the analysis.

5.4.8. Docetaxel dosing

The difference in incidence of docetaxel dose reduction and/or docetaxel dose withheld in Cycle 2 due to Neutropenia in Cycle 1 will be tested using Barnard's Test. The Method will also be used to construct point estimates and confidence intervals for difference. A 95% confidence interval of proportion with docetaxel dose reduction/withheld will be constructed by using the Clopper-Pearson method.

The difference of the number of docetaxel doses over all cycles of the DP Arm to the D Arm will be evaluated by using the Wilcoxon Rank Sum test. Hodges-Lehmann Estimation will also be used to construct point estimates and confidence intervals. The NPAR1WAY procedure in SAS version 9.4 or later will be used for the analysis.

The proportion of patients who received docetaxel more than 8 cycles, the proportion of patients who received docetaxel more than 10 cycles of the DP Arm to the D Arm will be summarized with descriptive statistics (N, percent) and analyzed for treatment differences using Barnard's test. The Method will also be used to construct point estimates and confidence intervals for difference. 95% confidence interval of proportions will be constructed by using the Clopper-Pearson method.

5.4.9. Population Pharmacokinetic Analysis

Population pharmacokinetic analyses will be conducted to evaluate the effect of intrinsic and extrinsic factors on the PK of plinabulin and its active metabolite(s), if identified, in humans. Intrinsic factors such as gender, age, hepatic or renal impairment, and race and/or ethnicity and extrinsic factors such as concomitant drugs, herbal products will be assessed in relation to drug exposure according to FDA Guidance for Industry, "Population Pharmacokinetics,"

Further details of PK analysis will be presented in the PK analysis plan.

6. REFERENCES

Blayney DW, McGuire BW, Cruickshank SE, Johnson DH. Increasing chemotherapy dose density and intensity: phase I trials in non-small cell lung cancer and non-Hodgkin's lymphoma. Oncologist. 2005;10(2):138-49.

Chow SC, Shao J, Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH Harmonised Tripartite Guideline, Statistical Principles for Clinical Trials (E9), 5 February 1998.

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH Harmonised Tripartite Guideline, Structure and Content of Clinical Study Reports (E3), 30 November 1995.

International Federation of Pharmaceutical Manufacturers and Associations. Medical Dictionary for Regulatory Activities (MedDRA). Version 14.0 Reston, Virginia, USA; 2008.

Jennison C, Turnbull BW. 2000. Group Sequential Methods with Applications to Clinical Trials. Chapman & Hall/CRC. New York.

Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika 1983;70: 659-63.

Mehta CR, Senchaudhuri P. Conditional versus Unconditional Exact Tests for Comparing Two Binomials. 2003.

O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics. 1979;35(3):549-56.

Pocock SJ. Group sequential methods in the design and analysis of clinical trials. Biometrika 1977:64;191-9.

Proschan, M., Lan, K.K.G., Wittes, J.T. 2006. Statistical Monitoring of Clinical Trials. Springer. New York

Protocol No. BPI-2358-103, Amendment 8, dated Jan 19, 2019.

Reboussin DM, DeMets DL, Kim K, et al. 1992. Programs for computing group sequential boundaries using the Lan-DeMets method. Madison, Wisconsin: Department of Biostatistics, University of Wisconsin-Madison; Technical Report 60.

WHO Collaborating Center for International Drug Monitoring. WHO Drug Dictionary. June 2012 B Format edition. Uppsala, Sweden; 2008.

7. APPENDIX

Appendix A: Stratification variables

Stratified randomization using blocks will be applied to randomize at a 1:1 ratio to receive docetaxel + plinabulin (DP Arm) or docetaxel + placebo (D5W) (D Arm). Stratification variables include geographic region, ECOG status; and prior lines of therapy:

- 1st-line platinum containing therapy (platinum combined with anything except docetaxel);
- 1st-line platinum containing therapy and 2nd-line PD-1 or PD-L1 inhibitor;
- 1st-line PD-1 or PD-L1 inhibitor and 2nd-line platinum containing therapy; or any other 1st or 2nd-line therapies (one of these 2 lines must have been a platinumbased regimen).

First- or second-line must contain a platinum-containing regimen. The first- or second-line can be any standard of care regimen including immunotherapy with a PD-1/PD--L1 checkpoint inhibitor. No prior docetaxel therapy for any stage of NSCLC is allowed.

In summary, there are four possible first- and second-line therapy combinations that patients may have received prior to study entry. These four combinations for each region-by-ECOG strata will yield 16 individual strata (as listed below) in which patients will be randomized in a 1:1 manner, using randomized blocks.

1. 1st-line =Platinum Containing Therapy, Region=China, and ECOG=0-1;
2. 1st-line =Platinum Containing Therapy, Region=China, and ECOG=2;
3. 1st-line =Platinum Containing Therapy, Region=RoW, and ECOG=0-1;
4. 1st-line =Platinum Containing Therapy, Region=RoW, and ECOG=2;
5. 1st-line=Platinum Containing Therapy + 2nd-line PD-1/PD-L1, Region=China, and ECOG=0-1;
6. 1st-line=Platinum Containing Therapy + 2nd-line PD-1/PD-L1, Region=China, and ECOG=2;
7. 1st-line=Platinum Containing Therapy + 2nd-line PD-1/PD-L1, Region=RoW, and ECOG=0-1;
8. 1st-line=Platinum Containing Therapy + 2nd-line PD-1/PD-L1, Region=RoW, and ECOG=2;
9. 1st-line=PD-1/PD-L1 inhibitor + 2nd-line Platinum Containing Therapy, Region=China, and ECOG=0-1;
10. 1st-line=PD-1/PD-L1 inhibitor + 2nd-line Platinum Containing Therapy, Region=China, and ECOG=2;
11. 1st-line=PD-1/PD-L1 inhibitor + 2nd-line Platinum Containing Therapy, Region=RoW, and ECOG=0-1;
12. 1st-line=PD-1/PD-L1 inhibitor + 2nd-line Platinum Containing Therapy, Region=RoW, and ECOG=2;
13. Any 1st or 2nd line treatment that does not contain a PD-1/PD-L1 inhibitor, Region=China, and ECOG=0-1;

14. Any 1st or 2nd line treatment that does not contain a PD-1/PD-L1 inhibitor, Region=China, and ECOG=2;
15. Any 1st or 2nd line treatment that does not contain a PD-1/PD-L1 inhibitor, Region=RoW, and ECOG=0-1;
16. Any 1st or 2nd line treatment that does not contain a PD-1/PD-L1 inhibitor, Region=RoW, and ECOG=2;