Document Cover Page

Official Title: A Phase 2/3, Multicenter, Randomized, Double Blind Study to Evaluate Duration of Severe Neutropenia with Plinabulin Versus Pegfilgrastim in Patients with Solid Tumors Receiving Docetaxel Myelosuppressive Chemotherapy

NCT Number: NCT03102606

Document Date: 19 August 2019

STATISTICAL ANALYSIS PLAN

PHASE 3 DATA

A Phase 2/3, Multicenter, Randomized, Double Blind, Study to Evaluate Duration of Severe Neutropenia with Plinabulin Versus Pegfilgrastim in Patients with Solid Tumors Receiving Docetaxel Myelosuppressive Chemotherapy

Sponsor: BeyondSpring Pharmaceuticals, Inc.

Protocol Number: BPI-2358-105

Phase: Phase 2/3

Author: Shuxin Yin, Ph.D.
Principle Biostatistician
BeyondSpring Pharmaceuticals, Inc.

Date: 19 August 2019

Version: 3.0

Signature Page

A Phase 2/3, Multicenter, Randomized, Double Blind, Study to Evaluate Duration of Severe Neutropenia with Plinabulin Versus Pegfilgrastim in Patients with Solid Tumors Receiving Docetaxel Myelosuppressive Chemotherapy

Protocol Number:	BPI-2358-105
Version:	3.0
Date:	19 August 2019

Statistical Analysis Plan Approval

Prepared by: _____ 2019-08-19

Shuxin Yin, Ph.D.
Principle Biostatistician
BeyondSpring Pharmaceuticals, Inc.
Date

Reviewed by: _____ 2019-08-19

Stephan Ogenstad, Ph.D.
Expert Biostatistician
Statogen Consulting LLC
Date

Approved by: _____ 2019-08-19

Ramon Mohanlal, M.D., Ph.D.
Chief Medical Officer
BeyondSpring Pharmaceuticals, Inc.
Date

Table of Contents

List of Abbreviations

AE	Adverse event
ALT	Alanine aminotransferase
ANC	Absolute neutrophil count
AST	Aspartate aminotransferase
ATC	Anatomical therapeutic chemical
AUC	Area under the curve
BOCF	Baseline observation carried forward
CI	Confidence intervals
CRO	Contract research organization
CSR	Clinical Study Report
CTCAE	Common Terminology Criteria for Adverse Events
DMSN	Duration of moderate and severe neutropenia
DSMB	Data safety monitoring board
DSN	Duration of severe neutropenia
DVM	Data validation manual
ECG	Electrocardiogram
eCRF	Electronic case report form
EORTC	European Organization for Research and Treatment of Cancer
FACS	Fluorescence-activated cell sorting
FDA	Food and Drug Administration
FN	Febrile neutropenia
G-CSF	Granulocyte colony-stimulating factor
HRPC	Hormone refractory prostate cancer
ICH	International Council on Harmonisation
ITT	Intent-to-treat
IV	Intravenous
LMR	Lymphocyte-to-Monocyte ratio
LOCF	Last observation carried forward
MCV	Mean corpuscular volume
MedDRA	Medical Dictionary for Regulatory Activities
NBR	Negative Binomial Regression
NCI	National Cancer Institute
NLR	Neutrophil-to-lymphocyte ratio
PDF	Portable document format
PLR	Platelet-to-Lymphocyte ratio
PP	Per-protocol
RTF	Rich Text Format
SAP	Statistical analysis plan
SOP	Standard operating procedures
SRS	Study report specification
TAC	Docetaxel, doxorubicin, and cyclophosphamide

1. Introduction

The purpose of this statistical analysis plan (SAP) is to describe the planned analyses and reporting for the Phase 3 portion of protocol BPI-2358-105 Protocol Amendment 6, dated 5 August 2019.

This SAP is written with due consideration of the recommendations outlined in the most recent International Conference on Harmonization (ICH) E9 Guideline entitled Guidance for Industry: Statistical Principles for Clinical Trials and the most recent ICH E3 Guideline, entitled Guidance for Industry: Structure and Content of Clinical Study Reports (CSR).

This SAP describes the data that will be analyzed and the subject characteristics, efficacy, and safety assessments that will be evaluated. This SAP provides details of the specific statistical methods that will be used. The statistical analysis methods presented in this document will supersede the statistical analysis methods described in the clinical protocol. If additional analyses are required to supplement the planned analyses described in this SAP, they may be completed and will be identified in the CSR.

2. Study Objectives

2.1 Primary Efficacy Objective (Cycle 1)

- To assess DSN in treatment Cycle 1 in patients with advanced or metastatic breast cancer, who have failed < 5 prior lines of chemotherapy; locally advanced or metastatic non-small cell lung cancer (NSCLC) after platinum therapy failure; or hormone-refractory (androgen-independent) metastatic prostate cancer (HRPC) treated with docetaxel (75 mg/m^2) + plinabulin (40 mg) (Arm 2)versus docetaxel (75 mg/m^2) + pegfilgrastim (6 mg) (Arm 1). ANC will be assessed at baseline (prior to Cycle 1 docetaxel dose) and during Cycle 1 on Days 1, 2, 6, 7, 8, 9, 10, and 15. Blood draws for ANC will be taken at approximately the same time as the time of the predose sample on Day 1 and will be taken by preference in the morning. DSN should be calculated as the number of consecutive days from the first day when a patient's ANC is below 0.5 x l0^9/L until the patient reaches an ANC> 0.5 x 10^9/L, in Cycle 1. For patients who do not experience any severe neutropenia in Cycle 1, the DSN is set to 0. For patient's experiencing several episodes, the number of days of DSN will be summed up.

2.2 Secondary Efficacy Objectives (Cycle 1 and Cycles 1 to 4)

- Platelet count in Cycle 1: maximum decrease from baseline (prior to Cycle 1 docetaxel dose)
- The proportion of patients with neutrophil-to-lymphocyte ratio (NLR) > 5 after Day 7 though Day 15 in Cycle 1
- The area under the concentration-time curve (AUC) using the trapezoidal quadrature method for bone pain, from Day 1 through Day 8 (pegfilgrastim will be administered on Day 2) in Cycle 1, based on the pain score from the patient bone pain scale
- Change in estimated mean bone pain score from pre-dose Day 1 through Day 8 in Cycle 1
- The proportion of patients with thrombocytopenia (all grade) in Cycle 1 to 4

2.3 Exploratory Efficacy Objectives (Cycle 1):

- To evaluate the proportion of patients in Cycle 1 with:
    1. Thrombocytopenia (all grade)
    2. Grade 4 neutropenia (ANC < 0.5 × 10^9/L)
    3. Grade 3 neutropenia (ANC < 1 × 10^9/L and ANC >=0.5× 10^9/L)
    4. Grade 3 and Grade 4 neutropenia
    5. Bands > 0 after Day 7 through Day 15
    6. Promyelocytes plus myelocytes > 0 after Day 7 through Day 15
    7. Lymphocyte-to-monocyte ratio (LMR) < 3.2 after Day 7 through Day 15
    8. Platelet-to-lymphocyte ratio (PLR) > 200 after Day 7 through Day 15
    9. At least 1 day of bone pain
    10. At least 2 days of bone pain
    11. At least 3 days of bone pain
    12. At least 4 days of bone pain
    13. At least 5 days of bone pain
    14. At least 6 days of bone pain
    15. At least 7 days of bone pain
    16. At least 8 days of bone pain
- To compare the proportion of patients who needed bone pain medication (defined as any medication reported on pain medication assessment from Day 1 through Day 8 in Cycle 1)

- To compare the time (in days) to the first use of bone pain medication between the treatment groups
- To assess DSN in Cycle 1 in patients with locally advanced or metastatic NSCLC after platinum therapy failure
- To assess DSN in Cycle 1 in patients with HRPC
- To assess DSN in Cycle 1 in patients with advanced or metastatic breast cancer who have failed < 5 prior lines of chemotherapy
- To assess DSN in Cycle 1 in patients with locally advanced or metastatic NSCLC after platinum therapy failure or HRPC
- Platelet count at least 30% change from baseline at any time during Cycle 1
- ANC nadir during Cycle 1
- For selected countries only: to investigate the following cytokine panel: IL-1beta, IL-6, IL-12p70, IL-12p40, IL-17A, IL-23, G-CSF, GM-CSF, IFN-alpha, IFN-gamma, TNF-alpha, IL-2, FLT-3 ligand, and IL-8

2.4 Exploratory Efficacy Objectives (Cycles 1 to 4):

- To evaluate the proportion of patients in Cycles 1 to 4 with:
    1. Febrile neutropenia (FN) (ANC $<1.0 \times 10^9$/L AND a single temperature of >38.3°C or a sustained temperature of ≥38°C for more than 1 hour)
    2. Grade 4 neutropenia
- To evaluate the following healthcare utilization objectives in Cycle 1 to 4:
    1. Incidence of 30-day rehospitalizations - all-cause
    2. Incidence of all-cause hospitalizations
    3. Duration of all-cause hospitalizations
    4. Incidence of all-cause emergency room (ER) visits
    5. Incidence of all-cause intensive care unit (ICU) stays
    6. Duration of all-cause ICU stays
    7. Incidence of all-cause docetaxel dose delay (≥7 days), dose reduction (≤85%), or dose discontinuation
    8. Incidence of platelet transfusions
    9. Incidence of antibiotics use

- To collect data on disease progression in Cycle 1 to 4
- To evaluate the following QoL objective in Cycle 1 to 4:
    1. Health-related Quality of Life (QoL) questionnaire evaluated with EORTC QLQ-C30 Item 30,
    2. Health-related Quality of Life (QoL) questionnaire evaluated with EORTC QLQ-C30 symptom summary score,
    3. EQ-5D-5L (“Your Health Today” Score),
    4. EQ-5D-5L health utility.

2.5 Safety Objectives

- Incidence, occurrence, and severity of adverse events (AEs)/serious adverse events (SAEs)
- Safety and tolerability (physical examination and safety laboratory assessments)

3. Study Endpoints and Analyses

3.1 Primary Efficacy Endpoint (Cycle 1)

DSN in treatment Cycle 1 in patients treated with docetaxel (75 mg/m^2) + plinabulin (40 mg) compared with patients treated with docetaxel (75 mg/m^2) + pegfilgrastim (6 mg). ANC will be assessed at baseline (prior to Cycle 1 docetaxel dose) and during Cycle 1 on Days 1, 2, 6, 7, 8, 9, 10, and 15. Blood draws for ANC will be taken approximately the same time as the time of the predose sample on Day 1, and will be taken by preference in the morning. The NBR model, on the integer number of days of severe neutropenia, will be used to analyze the DSN endpoint during the fixed time window outlined above, with the treatment arm as the only covariate. The method will also be used to construct point estimates and confidence intervals. The NBREG, GLM and TABSTAT procedures in Stata v11.0 or later will be used. The primary analysis will be conducted in the intent-to-treat analysis set and sensitivity analysis will be conducted in the safety analysis set.

3.2 Secondary Efficacy Endpoints (Cycle 1 and Cycles 1 to 4)

- Maximum decrease from baseline (prior to docetaxel dose) in platelet count for each patient in Cycle 1. The Wilcoxon Rank Sum test will be used. The method will also be used to construct point estimates and 2-sided confidence intervals. The NPAR1WAY procedure in SAS version 9.4 or later will be used for the analysis.
- Proportion of patients with NLR > 5 after Day 7 through Day 15 in Cycle 1. The Barnard's test will be used to evaluate the difference in proportions between the treatment arms. The method will also be used to construct point estimates and 2-sided confidence intervals. The FREQ procedure in SAS version 9.4 or later will be used for the analysis.
- AUC using the trapezoidal quadrature method for bone pain, from Day 1 through Day 8 in Cycle 1, based on the bone pain score from the patient bone pain scale. The Wilcoxon Rank Sum test will be used. The method 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 analysis will be conducted in the safety analysis set.
- Estimated mean pain score from the patient bone pain scale from pre-dose Day 1 through Day 8 in Cycle 1. A repeated measure mixed linear model with the pre-dose Day 1 value and treatment arm as covariates will be used to analyze this endpoint. 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. The analysis will be conducted in the safety analysis set.
- Proportion of patients with thrombocytopenia (all grade) in Cycles 1 to 4. The Barnard's test will be used to evaluate the difference in proportions between the treatment arms. The method will also be used to construct point estimates and confidence intervals. The FREQ procedure in SAS version 9.4 or later will be used for the analysis.

3.3 Exploratory Efficacy Endpoints (Cycle 1)

- Proportion of patients in Cycle 1 with:
    - Thrombocytopenia (all grade). Barnard's test will be used to evaluate the difference in proportions between the treatment arms. The method will also be used to construct point estimates and confidence intervals. The FREQ procedure in SAS version 9.4 or later will be used for the analysis.
    - Grade 3 and Grade 4 neutropenia. Barnard's test will be used to evaluate the difference in proportions between the treatment arms. The method will also be used to construct point estimates and confidence intervals. The FREQ procedure in SAS version 9.4 or later will be used for the analysis.
    - Grade 4 neutropenia. Barnard's test will be used to evaluate the difference in proportions between the treatment arms. The method will also be used to construct point estimates and confidence intervals. The FREQ procedure in SAS version 9.4 or later will be used for the analysis.
    - Grade 3 neutropenia. Barnard's test will be used to evaluate the difference in proportions between the treatment arms. The method will also be used to construct point estimates and confidence intervals. The FREQ procedure in SAS version 9.4 or later will be used for the analysis.
    - Bands > 0 after Day 7 through Day 15. Barnard's test will be used to evaluate the difference in proportions between the treatment arms. The method will also be used to construct point estimates and confidence intervals. The FREQ procedure in SAS version 9.4 or later will be used for the analysis.
    - Promyelocytes plus myelocytes >0 after Day 7 through Day 15. Barnard's test will be used to evaluate the difference in proportions between the treatment arms. The method will also be used to construct point estimates and confidence intervals. The FREQ procedure in SAS version 9.4 or later will be used for the analysis.
    - LMR < 3.2 after Day 7 through Day 15; time-course of the percentage of patients with LMR <3.2 in Cycle 1. Barnard's test will be used to evaluate the difference in proportions between the treatment arms. The method will also be used to construct point estimates and confidence intervals. The FREQ procedure in SAS version 9.4 or later will be used for the analysis.
    - PLR > 200 after Day 7 through Day 15; the time-course of the percentage of patients with PLR > 200 in Cycle 1. Barnard's test will be used to evaluate the difference in proportions between the treatment arms. The method will also be used to construct point estimates and confidence intervals. The FREQ procedure in SAS version 9.4 or later will be used for the analysis.
    - For the below bone pain measurements, Barnard's test will be used to evaluate the difference in proportions between the treatment arms in the safety analysis set. The method will also be used to construct point estimates and confidence intervals. The FREQ procedure in SAS version 9.4 or later will be used for the analysis:
        - At least 1 day of bone pain
        - At least 2 days of bone pain
        - At least 3 days of bone pain
        - At least 4 days of bone pain
        - At least 5 days of bone pain
        - At least 6 days of bone pain

  - At least 7 days of bone pain
  - At least 8 days of bone pain
- Proportion of patients in Cycle 1 who needed bone pain medication (defined as any medication reported on the pain medication assessment from Day 1 through Day 8). Barnard's test will be used to evaluate the difference in proportions between the treatment of patients who needed bone pain medication (defined as any medication reported on the pain medication assessment from Day 1 through Day 8 in Cycle 1).
- Time (in days) to the first use of bone pain medication. The log-rank test (LIFETEST procedure in SAS) will be used to compare time to the first use of bone pain medication between the treatment groups. Separate analyses will be made with respect to narcotic and non-narcotic analgesics.
- DSN in Cycle 1 in patients with locally advanced or metastatic NSCLC after platinum failure. The analysis will be done as per the primary endpoint.
- DSN in Cycle 1 in patients with HRPC. The analysis will be done as per the primary endpoint.
- DSN in Cycle 1 in patients with advanced or metastatic breast cancer who have failed < 5 prior lines of chemotherapy. The analysis will be done as per the primary endpoint.
- DSN in Cycle 1 in patients with locally advanced or metastatic NSCLC after platinum failure or HRPC. The analysis will be done as per the primary endpoint.
- Platelet count at least 30% change from baseline at any time during Cycle 1. Barnard's test will be used to evaluate the difference in proportions between the treatment arms. The method will also be used to construct point estimates and confidence intervals. The FREQ procedure in SAS version 9.4 or later will be used for the analysis..
- ANC nadir during Cycle 1. The Wilcoxon Rank Sum test will be used to analyze the endpoint. The method 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.
- For selected countries only: maximum change from baseline in the following cytokine panel: IL-1beta, IL-6, IL-12p70, IL-12p40, IL-17A, IL-23, G-CSF, GM-CSF, IFN-alpha, IFN-gamma, TNF-alpha, IL-2, FLT-3 ligand, and IL-8. The Wilcoxon Rank Sum test will be used to analyze the endpoint. The method 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.

3.4 Exploratory Efficacy Endpoints (Cycles 1 to 4)

- Proportion of patients in Cycles 1 to 4 with:
  - FN (ANC $<1.0 \times 10^9$/L AND a single temperature of >38.3°C or a sustained temperature of ≥38°C for more than 1 hour). Barnard's test will be used to evaluate the difference in proportions between the treatment arms. The method will also be used to construct point estimates and confidence intervals. The FREQ procedure in SAS version 9.4 or later will be used for the analysis
  - Grade 4 neutropenia. Barnard's test will be used to evaluate the difference in proportions between the treatment arms. The method will also be used to construct point estimates and confidence intervals. The FREQ procedure in SAS version 9.4 or later will be used for the analysis
- Healthcare utilization endpoints will be analyzed in the safety analysis set:
  - Incidence of 30-day rehospitalizations - all-cause. The NBR model will be

used with the treatment arm as the only covariate. An offset variable will be used to account for exposure time. The method will also be used to construct point estimates and confidence intervals. The NBREG, GLM and TABSTAT procedures in Stata v11.0 or later will be used.

- Incidence of all-cause hospitalizations. The NBR model will be used with the treatment arm as the only covariate. An offset variable will be used to account for exposure time. The method will also be used to construct point estimates and confidence intervals. The NBREG, GLM and TABSTAT procedures in Stata v11.0 or later will be used.
- Duration of all-cause hospitalizations. The Wilcoxon Rank Sum test will be used. The method 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.
- Incidence of all-cause ER visits. The NBR model will be used with the treatment arm as the only covariate. An offset variable will be used to account for exposure time. The method will also be used to construct point estimates and confidence intervals. The NBREG, GLM and TABSTAT procedures in Stata v11.0 or later will be used.
- Incidence of all-cause ICU stays. The NBR model will be used with the treatment arm as the only covariate. An offset variable will be used to account for exposure time. The method will also be used to construct point estimates and confidence intervals. The NBREG, GLM and TABSTAT procedures in Stata v11.0 or later will be used.
- Duration of all-cause ICU stays. The Wilcoxon Rank Sum test will be used. The method 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.
- Incidence of all-cause docetaxel dose delay and dose reduction. The NBR model will be used with the treatment arm as the only covariate. An offset variable will be used to account for exposure time. The method will also be used to construct point estimates and confidence intervals. The NBREG, GLM and TABSTAT procedures in Stata v11.0 or later will be used.:
  - Docetaxel dose delays > 7 days
  - Regimen switching
  - Treatment discontinuation
- Proportion of Relative dose intensity (RDI) ≤ 85% where the RDI is defined as

$$\text{RDI} = \frac{\text{Cumulative dose of administered docetaxel (mg/m}^2\text{)/ Number of weeks from Day1 Cycle 1 to Day 1 Cycle 4 plus 3 weeks}}{\text{Total calculated doses/12 weeks}}$$

  Barnard's test will be used to evaluate the difference in proportions between the treatment arms. The method will also be used to construct point estimates and confidence intervals. The FREQ procedure in SAS version 9.4 or later will be used for the analysis
- Incidence of transfusions due to thrombocytopenia. The NBR model will be used with the treatment arm as the only covariate. An offset variable will be used to account for exposure time. The method will also be used to construct point estimates and confidence intervals. The NBREG, GLM and

TABSTAT procedures in Stata v11.0 or later will be used

- Incidence of antibiotics use. The NBR model will be used with the treatment arm as the only covariate. An offset variable will be used to account for exposure time. The method will also be used to construct point estimates and confidence intervals. The NBREG, GLM and TABSTAT procedures in Stata v11.0 or later will be used.

- Data collection on disease progression. The log-rank test (LIFETEST procedure in SAS version 9.4 or later) will be used to compare time to disease progression between the treatment groups. The method will also be used to construct point estimates and confidence intervals.
- QoL endpoints will be analyzed in the safety analysis set:
  - Health-related QoL questionnaire evaluated with EORTC QLQ-C30 Item 30 (response to the question "How do you rate your overall quality of life during the past week") in Cycle 1 to 4 will be analyzed using a linear mixed model for repeated measures (MMRM) with the Cycle 1 Day 1 value and treatment arm as covariates. The method will also be used to construct point estimates and confidence intervals. The MIXED procedure in SAS v9.4 or later will be used for the analysis. The summary statistics in terms of counts, means, standard deviations, medians, minimums, and maximums for individual questions will be summarized in a tabular format.
  - Health-related QoL questionnaire evaluated with EORTC QLQ-C30 will also be analyzed with the summary scores. The EORTC QLQ-C30 scoring method combines the 30 questions from the questionnaire into 15 scores: global health state/QoL, functional scale (5 items), and symptoms scale (9 items) (Fayers et al, 2001). Three summary measures will be constructed: the QLQ-C30 Summary Score, which combined the symptom and functional scales (excluding the financial difficulties item), the symptom summary score (excluding the financial difficulties item), and the functional summary score (Hinz et al, 2012). The symptom summary score will be analyzed using a linear mixed model for repeated measures (MMRM) with the Cycle 1 Day 1 value and treatment arm as covariates. The method will also be used to construct point estimates and confidence intervals. The MIXED procedure in SAS v9.4 or later will be used for the analysis. The summary statistics in terms of counts, means, standard deviations, medians, minimums, and maximums for the 15 scores, the QLQ-C30 summary score, symptom score, and the functional summary score will be summarized in a tabular format.
  - The response to the question "Your Health Today" from EQ-5D-5L in Cycle 1 to 4 will be analyzed using linear mixed for model repeated measures (MMRM) with the Cycle 1 Day 1 value and treatment arm as covariates. The method will also be used to construct point estimates and confidence intervals. The MIXED procedure in SAS v9.4 or later will be used for the analysis. The summary statistics in terms of counts, means, standard deviations, medians, minimums, and maximums for the question "Your Health Today" will be summarized in a tabular format.
  - The EQ-5D-5L data will be converted into a health utility score using the interim EQ-5D-5L crosswalk value set for the United States (van Hout et al, 2012) unless a validated EQ-5D-5L value set for the United States becomes available. The EQ-5D-5L health utility score in Cycle 1 to 4 will be analyzed using a linear mixed model for repeated measures (MMRM) with the Cycle 1 Day 1 value and treatment arm as covariates. The method will also be used to construct point estimates and confidence intervals. The MIXED procedure in SAS v9.4 or later will be used for the analysis. The summary statistics in terms

of counts, means, standard deviations, medians, minimums, and maximums for the health utility score will be summarized in a tabular format.

3.5 Safety Endpoints

- Incidence, occurrence, and severity of AEs/SAEs. These endpoints will be presented descriptively in tables and listings.
- Safety and tolerability (physical examination and safety laboratory assessments). These endpoints will be presented descriptively in tables and listings.

3.6 Statistical Hypotheses

The null and alternative hypotheses for this non-inferiority trial design of duration of severe neutropenia (DSN) are as follows:

H_0: True difference ([arm 2: Docetaxel (75 mg/m^2) + plinabulin (40 mg) + placebo matching pegfilgrastim] minus arm 1: [Docetaxel (75 mg/m^2) + pegfilgrastim (6 mg) + placebo matching plinabulin]) in the mean DSN $\geqslant$ 0. 65 days

H_1: True difference (arm 2 minus arm 1) in the mean DSN < 0.65 days

Plinabulin will be considered to be demonstrated to be non-inferior to pegfilgrastim if in Cycle 1, the upper limit of the 2-sided 95% confidence interval for the true difference in the mean duration of Grade 4 neutropenia < 0.65 days. If non-inferiority is demonstrated, then the upper limit of the 2-sided 95% confidence interval will be compared to 0. If the upper limit < 0, the superiority of plinabulin to pegfilgrastim will be concluded. Specifics of the statistical tests are provided in Section 14.

4. Study Design and Procedures

4.1 General Study Design

This is a multicenter, randomized study with an open-label Phase 2 portion and a double-blind Phase 3 portion. This SAP pertains to the Phase 3 portion of the study only.

Phase 3 will not begin until RP3D has been determined based on the phase 2 PK/PD analysis as mentioned above; the RP3D will be the only plinabulin dose administered in Phase 3.

A fixed dose of 40 mg has been selected as the RP3D following the Phase 2 PK/PD analysis (details of dose selection are provided in Section 6.2.3 of protocol amendment 6).

Approximately 150 patients are planned to be enrolled in Phase 3 with one of the following diagnosis: advanced or metastatic breast cancer, who have failed < 5 prior lines of chemotherapy; advanced or metastatic NSCLC who have previously received platinum-based therapy and have either disease progression, are intolerant of platinum-based therapy or, in the opinion of the investigator, would benefit from docetaxel chemotherapy; or hormone-refractory (androgen-independent) metastatic prostate cancer. Each eligible patient will be stratified according to his or her tumor type (breast cancer, NSCLC or HRPC) and region (Asia, non-Asia). Patients will be randomly assigned within each stratum (diagnosis) with equal probability (1:1 ratio) or 75:75, with the arm designation and planned intervention as follows:

Arm 1: Docetaxel (75 mg/m^2) + pegfilgrastim (6 mg) + placebo matching plinabulin

Arm 2: Docetaxel (75 mg/m^2) + plinabulin (40 mg) + placebo matching pegfilgrastim

Data from all patients receiving the RP3D plinabulin dose in Phase 2 and Phase 3 will not be pooled for assessing the primary and secondary study endpoints but analyzed separately.

Treatments Administered:

Both Phase 2 and Phase 3, Cycles 1 to 4, will consist of docetaxel 75 mg/m^2 administered by intravenous (IV) infusion on Day 1 over 60 minutes (±5 minutes) every 21 days. In the phase 2 portion, on Day 1 of each cycle, 1.5 hours (± 10 minutes) after the start time of docetaxel infusion (i.e., approximately 30 minutes after the end of docetaxel infusion), patients assigned to a plinabulin arm (arms 2-4) will get a single intravenous infusion of plinabulin at their assigned dose over 30 minutes (± 5 minutes). Thus, the wait time between the end of docetaxel infusion and start of the plinabulin infusion is approximately 30 minutes. On Day 2 of each cycle, ≥24 hours after completing chemotherapy, patients assigned to pegfilgrastim (arm 1) will receive a single dose of pegfilgrastim (6 mg) (subcutaneous injection).

In the phase 3 portion, on Day 1 of each cycle, 1.5 hours (± 10 minutes) after the start time of docetaxel infusion (i.e., approximately 30 minutes after the end of docetaxel infusion), patients will get a single dose of plinabulin or placebo intravenously over 30 minutes (± 5 minutes). On Day 2 of each cycle, ≥24 hours after completing chemotherapy, patients will receive a single dose of pegfilgrastim (6 mg) or placebo (subcutaneous injection).

If a chemotherapy cycle is delayed by more than 3 weeks, the patient will be withdrawn from the study. If a critical AE (refer to Section 10.5 of protocol amendment 6) occurs during the cycle, the dosage of docetaxel may be reduced by 20% in the next cycle. Only one docetaxel dose reduction is allowed (refer to Section 11.5 of protocol amendment 6)**Error! Reference source not found.** No dose reductions are allowed with plinabulin or pegfilgrastim.

All patients, including patients who withdraw from the study early, will complete a safety follow-up visit 30 days (± 2 days) after the last dose of study drug. If in the opinion of the investigator, the patient will benefit from more than 4 cycles of docetaxel and open-label pegfilgrastim, then the fifth cycle will not start until completion of the safety follow-up visit (in this instance, the safety follow-up visit will be Cycle 4 Day 21). Follow-up visits will be required to monitor for ongoing treatment-related AEs. All

patients experiencing drug-related toxicities of ≥ Grade 2 at the End of Treatment visit should be followed-up at least monthly until the AE(s) resolves to ≤ Grade 1, the event is considered to be chronic, or the patient receives other anti-cancer therapy. The method of follow-up assessment will be at the Investigator's discretion (for example, patient site visit or telephone call). All deaths which occur within 30 days of study drug administration regardless of relationship to the study drug must be reported to the Sponsor immediately and within 24 hours of becoming aware of the event.

Laboratory test results (hematology and serum chemistry) will be collected via a central laboratory. Safety laboratory tests are required prior to treatment on Day 1 of each cycle and can be collected by a local laboratory; however, all other scheduled blood samples as per the schedule assessments and procedure table must also be obtained for central laboratory assessment. Urinalysis will be performed at screening only. CD34+ counts will be established through a fluorescence-activated cell sorting (FACS) method as described in Table 6 and Table 7 of protocol amendment 6.

5. Study Treatments

A planned 75 patients per arm with advanced or metastatic breast cancer, who have failed < 5 prior lines of chemotherapy; advanced or metastatic NSCLC who have previously received platinum-based therapy and have either disease progression, are intolerant of platinum-based therapy, or in the opinion of the investigator, would benefit from docetaxel chemotherapy; or hormone-refractory [androgen independent] metastatic prostate cancer:

- Arm 1: Docetaxel (75 mg/m^2) + pegfilgrastim (6 mg) + placebo matching plinabulin
- Arm 2: Docetaxel (75 mg/m^2) + plinabulin (40 mg) + placebo matching pegfilgrastim

All patients with HRPC also are given prednisone 5 mg orally twice daily continuously in addition to docetaxel (see Taxotere® (Prescribing Information)) as well as other assigned study drugs. In the phase 3 portion, on Day 1 of each cycle, 1.5 hours (± 10 minutes) after the start time of docetaxel infusion (i.e., approximately 30 minutes after the end of docetaxel infusion), patients will get a single dose of plinabulin or placebo intravenously over 30 minutes (± 5 minutes). On Day 2 of each cycle, ⩾24 hours after completing chemotherapy, patients will receive a single dose of pegfilgrastim (6 mg) or placebo (subcutaneous injection).

6. Sample Size and Power Considerations

Approximately 150 patients are planned to be enrolled with 1 of the following diagnoses: advanced or metastatic breast cancer, NSCLC, or HRPC. A sample size of 75 patients in each of the treatment arms: docetaxel + plinabulin (40 mg) versus docetaxel + pegfilgrastim, with matching placebos achieve at least a 90% power to reject the null hypothesis of 0.65 days of inferiority in DSN between the treatment means with standard deviation of 0.75, at an overall two-sided significance level (alpha) of 0.05, using a two-sample zero-inflated Poisson model and an O'Brien-Fleming spending function to account for the interim analysis at 2/3rds (66.7%) of information.

The software PASS version 15.0.1 has been used for the calculations referencing Chow et al. 2003, Lan and DeMets 1983, O'Brien and Fleming 1979.

Timing of Statistical Analysis

	Number of patients in Phase 3 (approximate)	
Arm 1: Docetaxel + Pegfilgrastim 6 mg	50	75
Arm 2: Docetaxel + Plinabulin (40 mg)	50	75
	Interim Analysis	**Final Analysis**

7. Data Preparation

Data management procedures, including database design, selection of the data dictionary, and coding of all adverse events and medications, will be performed by Statistics & Data Corporation (SDC). All reported study data will be recorded on the eCRFs supplied by SDC using Medidata Rave®. Clinical personnel at the study center and CRO are responsible for ensuring that the protocol is followed and that the eCRFs are properly completed.

After data are entered into the clinical study database, electronic edit checks will be performed, including checks for missing data, out of range values, discrepancies within and across visits, and cross-checks between different data tables. All data validation specifications and procedures are detailed in the Data Validation Manual (DVM), and manual data checks are documented in the Study Report Specifications (SRS). When the database has been declared to be complete and accurate, the database will be locked, and treatment codes unmasked. Any changes to the database after that time can only be made with the approval of the Sponsor in consultation with SDC.

All final analyses outlined in this document will be performed after:

- All data management requirements are met according to SDC's Standard Operating Procedures (SOP), including a performance of edit and validation checks, documentation and resolution of data queries, and database lock with written authorization provided by appropriate SDC and Sponsor personnel;
- All protocol deviations have been classified as major or minor and the per-protocol (PP) population has been determined; and
- The treatment codes have been unmasked

8. Analysis Sets

8.1 Intent-to-Treat (ITT) Analysis Set

The intent-to-treat (ITT) analysis set for Phase 3 is comprised of all Phase 3 patients that have been randomized in the study.

The analysis of all efficacy endpoints, unless noted otherwise, will be conducted on the intent-to-treat analysis set.

8.2 Safety Analysis Set

The safety analysis set is comprised of all Phase 3 patients that have been randomized in the study and have received at least one dose of study medication.

9. Efficacy Analyses Considerations

9.1 Primary efficacy analysis considerations

DSN was not measured in the previous Phase 2 study ([Study NPI-2358-101] in which patients received treatment with either plinabulin (30 mg/m^2 [n=50] or 20 mg/m^2 [n=40] + docetaxel or docetaxel alone [n=73]). DSN was obtained using the following methods (described below) for the the generation of ANC data and the observed neutrophil values on Day 8/Cycle 1 in the Phase 2 study. Day 8 neutrophil values were shown to approximately coincide with the nadir of ANC after docetaxel treatment (Blackwell et al., 2015). The study will assume that the shape of the time/neutrophil recovery curve in plinabulin-treated patients is indistinguishable from the time/neutrophil recovery curve for filgrastim and its biosimilars.

In a study with filgrastim and its biosimilar, the time-course of ANC in Cycle 1 for the Per Protocol dataset was published by Blackwell et al., 2015. Mean values and standard deviations of ANC during the 21-day follow-up period were readily available. This information was used to write a computer simulation program that would generate random ANC data that asymptotically has the same means and standard deviations for the 21-day follow-up period as the publication. The simulation would then also generate the projected number of days with severe neutropenia (i.e., the DSN).

Deming regression (Deming, 1943) was used to calculate the linear relationship between simulated nadir and DSN. The rank correlation between simulated nadir and DSN was used to calculate the DSN with plinabulin (+ docetaxel) and docetaxel alone. In the Phase 2 study, ANCs were obtained on Day 8, which approximately coincides with the time that the neutrophil nadir occurs after docetaxel administration. The observed Day 8 neutrophil (nadir) values were computed into the linear relationship (Deming regression), mentioned above to calculate DSN for each patient. Using these methods, calculated mean DSN was 0.065 days for the plinabulin+ docetaxel arm, and 1.076 days for the docetaxel alone. Based on published data with filgrastim in patients receiving docetaxel (Alexopoulos K et al., 1999), the assumption is that Grade 4 neutropenia in Cycle 1 would occur in a 2 times higher frequency with G-CSF+docetaxel versus plinabulin+docetaxel, resulting in a presumed mean DSN of 0.13 days for the G-CSF+ docetaxel combination.

This non-inferiority trial design will utilize a difference (arm 2 minus arm 1) of 0.65 days (non-inferiority margin) in DSN in Cycle 1 as the largest acceptable difference between plinabulin and pegfilgrastim. The non-inferiority test will evaluate the null hypothesis H_0: true difference (arm 2 minus arm 1) ≥ 0.65 against the alternative hypothesis H_1: true difference (arm 2 minus arm 1) < 0.65. Plinabulin will be considered non-inferior to pegfilgrastim if, in Cycle 1, the upper limit of the 2-sided 95% confidence interval for the true difference in the mean duration of Grade 4 neutropenia was < 0.65 days. A sample size of patients was based on sample size considerations as outlined.

Data suggest (http://www.neulastahcp.com/risk/duration-of-severe-neutropenia/) that FN is correlated with DSN. The frequency of FN with docetaxel monotherapy (100 mg/m2) + G-CSF was reported to be 1% in Cycle 1. FN frequency in Cycle 1 with docetaxel combined with doxorubicin and G-CSF was ~ 3 % (Aarts M, et al. 2013), which would translate into a DSN of 1 day according to Holmes FA, et al, 2002. Based on these data, it is assumed that the median DSN for docetaxel monotherapy + G-CSF will be approximately 1 day.

The frequency of FN with docetaxel monotherapy (without G-CSF) has been reported to be 11% in Cycle 1 (17% over all cycles) docetaxel dose of 100 mg/m^2 (Vogel et al, 2005) and 19.8% over all cycles at a lower docetaxel dose of 60 mg/m^2 (Yoh K et al., 2016). Hanna N et al., 2004 reported an FN percentage of 12.7% with 75 mg/m^2 docetaxel. Based on this range of FN, the relationship established by Meza et al., 2002 between FN and DSN, we make the assumption that, with docetaxel monotherapy at a dose of 75 mg/m^2 without G-CSF, the median DSN is estimated to be 4-5 days.

In the Zarxio® briefing document, 2015, the margin was selected based on the fact that Taxotere/Adriamycin/cyclophosphamide (TAC) chemotherapy is known to induce a median DSN of 7 days in breast cancer patients receiving no G-CSF treatment (Nabholtz, 2001), while G-CSF treatment reduces the mean DSN for this chemotherapy to 1.4 days (95% CI: 1.07 - 1.69) as shown in pegfilgrastim (Neulasta®) Study 20020778 (Kaufman et al., 2004). Based on this a non-inferiority limit of 1 day was derived.

As an extension of this reasoning, it is argued for our study, a non-inferiority margin of 0.65 would be reasonable and correspond to approximately a median of 4.5 days of DSN, as a ratio of 1 day to 7 days of DSN in the Zarxio® briefing document, 2015.

A non-inferiority margin of 0.65 days can also be justified because a difference of 0.65 days is not considered to be clinically meaningful.

10. General Statistical Considerations

10.1 Definition of Baseline

Baseline measures are defined as the last non-missing measure prior to receipt of randomized study treatment.

10.2 Data Analysis Conventions

All data analyses at the end of the trial will be performed by SDC. Statistical analyses will be reported using summary tables, figures, and data listings. Continuous variables will be summarized with counts, means, standard deviations, medians, minimums, and maximums. Categorical variables will be summarized by counts and by the percentage of patients. Confidence intervals will be presented in the summary for continuous and/or categorical variables as deemed appropriate.

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

Individual patient data obtained from the case report forms (CRF), electrocardiogram (ECG), core laboratory, 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 later. Tables, listings, and figures will be presented in Portable Document Format (PDF), and tables will be presented in Rich Text Format (RTF) as well. Upon completion, all SAS programs will be validated by an independent programmer. In addition, all program output will undergo a statistical review, a quality control specialist review and a 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.

10.3 Missing or Inconclusive Data Handling

Missing data will not be imputed for the primary and key secondary endpoint analyses but will be used for sensitivity analyses. ANC values that are too low to be calculated, and which have the corresponding grade 3 or 4 WBC values that are typically associated with Grade 4 Neutropenia will not be considered missing data. ANC values that are too low to be calculated, and which have the corresponding grade 3 or 4 WBC values that are typically associated with Grade 4 Neutropenia will not be considered missing data. Sensitivity analyses will be performed for DSN using multiple-imputation models (Mehrotra et 2017) to examine the potential impact of the missing ANC data. Baseline characteristics which should be considered in the imputation model include but are not restricted to the baseline values of the respective efficacy variable, the treatment group, and sex.

Missing pain score data: In the case in which analgesic medication was taken during the treatment period, as sensitivity analyses, the pain scores will be imputed using the last observation carried forward (LOCF), Worst Observation Carried Forward (WOCF), and Baseline Observation Carried Forward (BOCF) methods. AUC using the trapezoidal quadrature method for bone pain will not be imputed.

10.4 Adjustments for Multiplicity

The study has two components to multiplicity. The first component is an interim analysis. The study design is group sequential with one interim analysis (after 50 patients in each treatment arm have completed at least one cycle in each of the treatment arms docetaxel + plinabulin [40 mg] versus docetaxel + pegfilgrastim, with matching placebos) and one final analysis at the completion of the

study. These results assume that two sequential tests are made using the O'Brien-Fleming spending function to determine the test boundaries.

The second multiplicity component of the study is initially testing the primary endpoint to determine non-inferiority of plinabulin to pegfilgrastim. If non-inferiority is demonstrated, then the superiority of plinabulin to pegfilgrastim will be tested. As this is a hierarchical testing procedure, no additional alpha adjustment is necessary.

At the interim analysis, if non-inferiority is demonstrated, then superiority will also be tested. If the plinabulin treatment is concluded to be superior to the pegfilgrastim treatment, with respect to DSN, then the study will be stopped. Additionally, since the design allows for stopping for inferiority, it might be decided on that occasion that the study will be stopped.

11. Disposition of Subjects

The subject disposition will be presented in terms of the numbers and percentages of subjects who were randomized, completed the study, and discontinued from the study. Subjects who are not discontinued from the study will be considered study completers. The disposition will be summarized by treatment group and for all subjects.

The number and percentage of subjects completing the study or who prematurely discontinued from the study and the reasons for study discontinuation will be summarized by treatment group for all randomized subjects. A subject listing will be provided that includes the date of completion or withdrawal, and reason for premature study discontinuation.

The number and percentage of subjects with protocol deviations will be summarized by treatment group for all randomized subjects. A subject listing will be provided that includes the date of the deviation, the deviation description and the classification of whether the deviation was judged to be major or minor.

In addition, subject listings will be provided that include informed consent date, randomization date, randomization stratum, inclusion and exclusion criteria violations.

12. Demographic and Other Baseline Variables

12.1 Demographic Variables

A summary of age, sex, race, ethnicity and subject's cancer type will be presented using appropriate descriptive statistics by each treatment arm and overall total. The categorical (discrete) variables will be summarized using counts and percentages. The continuous variables will be summarized using counts, mean, median, standard deviation, and range (minimum, maximum). These summaries will include patients in the ITT analysis set.

All demographic and baseline characteristics will be listed by the study-center and subject number.

12.2 Baseline Characteristics

Cancer diagnosis, cancer staging, sample collection, bone marrow involvement, and disease progression will be summarized by using frequency counts and percentages for the ITT analysis set.

The child-bearing potential and any other assessments that are done for the purpose of eligibility for inclusion into the study (physical examination, vital signs, hematology and blood chemistry, urinalysis, pregnancy test, and ECG) will also be summarized using appropriate descriptive statistics by each treatment arm and overall total for the ITT analysis set.

13. Medical History and Concomitant Medications

13.1 Medical History

The medical history will be summarized using discrete summary statistics and presented by treatment group at the subject and event-level by system organ class (SOC) and preferred term (PT). If a subject reports the same PT multiple times within the same SOC, that PT will only be reported once within that SOC. As with the PT, if a subject reports multiple conditions within the same SOC, then that SOC will only be reported once. The medical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA), version 20.1. The summaries will be based on the safety analysis set.

Listings of medical history will also be generated.

13.2 Cancer History and Prior Cancer-Related Therapies

Prior cancer-specific surgery, radiotherapy, prior chemotherapy regimen, prior anti-cancer therapy for the investigational disease including drug name, dose, number of cycles completed and reason for therapy discontinuation and best overall response will be summarized using frequency counts and percentage.

13.3 Prior and Concomitant Therapies (Medication, Treatments, and Surgical Procedures)

All prior and concomitant medications, treatments, and surgical procedures will be listed using the safety population including generic name, route of administration, start date, stop date, dosage, indication, related AE or medical history, and reason for treatment. Prior and concomitant medications will be coded using the WHO Drug dictionary enhanced (WHO-DDE) + H B2, September 2017, to the appropriate Anatomical Therapeutic Chemical (ATC) classification and WHO generic term. Prior and concomitant treatments and surgical procedures will be coded using MedDRA, version 20.1.

Counts and percentages of prior and concomitant medications will be summarized using WHO-DDE ATC classification and preferred name. Summaries will be displayed by treatment group and for all subjects. Subjects with multiple medications in the same ATC class or preferred name will be counted only once for that respective ATC class or preferred name.

13.4 Study Drug Exposure

Study treatment exposure will be summarized in the safety population.

For each treatment arm for each product (docetaxel, pegfilgrastim, and plinabulin), the following will be summarized using descriptive statistics by study arm and overall:

- Duration of exposure calculated as (date of the last dose – date of the first dose) + 1 day.
- Number of cycles received per patient.
- Number of cycles with dose delay, dose adjustment, dose interruption.
- Reasons for dose delay from planned therapy.
- Reasons for dose interruption from planned therapy.
- Reasons for Reason Actual Volume Administered is not the same as Volume Prepared: from planned therapy.
-

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

14. SAS Codes for Efficacy Analyses

Recurrent events using NBR model:

A Negative Binomial Regression (NBR) model on the integer number of days of severe neutropenia will be used to analyze the DSN endpoint during the fixed time window outlined above, with the treatment arm as the only covariate. This method will also be used to construct point estimates and confidence intervals. The NBREG, GLM and TABSTAT procedures in Stata v11.0 or later will be used as follows:

```
USE "INDATA"
NBREG DSN TX01, NOLOG
GLM DSN TX01, NOLOG FAMILY(NB) EFORM IRLS DISP(`E(ALPHA)') SCALE(DEV)
   MARGINS, OVER(TX01)
TABSTAT DSN, BY(TX01) STATS (MEAN V SD SEMEAN N)
```

where

- *INDATA* is the name of the input dataset;
- *DSN* is the count of DSN in treatment Cycle 1;
- *TX01* is the name of the treatment group variable;

The above procedures will be used for other recurrent events which will be analyzed by using NBR model and offset will be added if indicated in section 3.

Proportions using Barnard's Test:

The FREQ procedure in SAS will be used for the analysis as following:

```
PROC FREQ DATA=INDATA;
   TABLES TREATMENT*OUTCOME/;
   WEIGHT COUNT;
    EXACT BARNARD;
RUN;
```

where

- *indata is the name of the input dataset*
- *treatment is the treatment group;*
- *outcome is possible outcome of corresponding endpoints*

Continuous variable using the Wilcoxon Rank Sum method:

The NPAR1WAY procedure in SAS will be used for the analysis.

```
PROC NPAR1WAY DATA=INDATA WILCOXON;
      CLASS TREATMENT;
      VAR OUTCOME;
RUN;
```

where

- *indata is the name of the input dataset*
- *treatment is the treatment group;*
- *outcome is the corresponding endpoint*

Time to event variable using log-rank test:

The LIFETEST procedure in SAS will be used for the analysis as following:

```
PROC LIFETEST DATA=INDATA;
  TIME TIME* STATUS (0);
```

```
  STRATA TREATMENT;
RUN;
```

where

- *indata is the name of the input dataset*
- *treatment is the treatment group;*
- *Time is the time to first use of bone pain medication or time to disease progression between the treatment groups;*
- *Status is the censoring indicator. A status of 1 indicates an event time, and a status of 0 indicates a censored time.*

Multiple measures variable using a linear mixed model for repeated measures (MMRM):

The MIXED procedure in SAS will be used for these analyses as follows:

```
PROC MIXED DATA=INDATA;
  CLASS TREATMENT SUBJID AVISITN;
  MODEL RESP= TREATMENT BASELINE;
  REPEATED AVISITN / SUBJECT = SUBJID TYPE=UN;
  LSMEANS TREATMENT / CL PDIFF;
RUN;
```

where

- *indata is the name of the input dataset*
- *treatment is the treatment group;*
- *subjid is the subject identifier*
- *avisitn is the visit identifier*

baseline is the pre-dose Day 1 value; and resp is the response variable, e.g., EORTC QLQ-C30 Item 30, EORTC QLQ-C30 symptom summary score, EQ-5D-5L response to the question "Your Health Today", or the EQ-5D-5L health utility.

15. Safety Analyses

All patients will be evaluable for safety analysis if they receive at least 1 dose of study drug. The safety data will be presented by study arm in individual listings and summary tables, including frequency tables for AEs and frequency and shift tables for laboratory variables. All AEs 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. AEs and SAEs will be reported in combined tables and SAEs will also be in tabulated in a separate table.

All safety information will be listed by the study-center and subject number.

15.1 Adverse Events

An AE is defined as any untoward medical occurrence associated with the use of an investigational product (IP) in humans, whether or not considered IP-related. Treatment-emergent AEs are those events that occur after the first administration of any study therapy through 30 days post last dose of any study therapy. AEs with unknown onset dates will be counted as TEAEs. All AEs will be assigned a toxicity grade from grade 1 to grade 5. Their relationship to IP will be classified as suspected (related, probable or possible relationship), or not suspected (unlikely, unrelated or N/A - Drug Not Given). TEAEs with a missing (or unknown) relationship to study treatment will assume the greatest relationship to study treatment. TEAEs with missing severity grades will be categorized as “missing” for the tabulation of TEAEs by severity. Documentation of AEs will include onset date, toxicity grade, action(s) taken, IP relationship, outcome, resolution date, and seriousness. All AEs will be coded using MedDRA classifications with reference to SOCs and PTs (MedDRA version 20.1).

An overall tabular summary of AEs will be presented that includes the number of events and the number and percentage of subjects who experienced at least one event, by treatment group and overall. This summary will also include breakdowns of AEs further categorized as SAEs, AEs by maximal severity, AEs leading to dose delays or discontinuation of the patients from the study and AEs resulting in death. AEs will be summarized using discrete summary statistics and presented by treatment group at the subject and event-level by SOC and PT using the Safety population. If a subject reports the same PT multiple times within the same SOC, that PT will only be reported once within that SOC. As with the PT, if a subject reports multiple conditions within the same SOC, that SOC will only be reported once. The occurrence of AEs suspected to be related to IP will also be tabulated by SOC and PT. All AEs will be presented in a listing.

15.2. Physical Examination

A physical examination will be conducted at screening, Day 1 in all cycles and at the end of treatment (EOT) visit. A comprehensive physical examination will include evaluations of the head, eyes, ears, nose, throat, neck, chest (including heart and lungs), abdomen, limbs, skin, and a complete neurological examination. A urogenital examination will only be required in the presence of clinical symptoms related to this region.

Height will be measured in centimeters once at the screening. Weight will be measured in kilograms at screening, prior to Day 1 dosing in all cycles and at the end of treatment (EOT) visit. For Cycle 1, bodyweight will also be measured on Days 2 and 6. The physical examination will be listed by visit and treatment arms, but no tabular summaries will be produced.

15.3 Vital Signs

Vital signs assessments will be conducted at screening, Day 1, 2, 6, 7, 8, 9,10 and 15 in Cycle 1, Day 1 and 8 in Cycle 2 to 4, end of treatment, early discontinuation, and 30-day safety follow-up, and include temperature, systolic blood pressure (SBP), diastolic blood pressure (DBP), heart rate, and respiratory rate.

Vital signs assessments and changes from baseline in vital signs assessments will be summarized by visit and treatment arm using quantitative summary statistics

15.4 Clinical Laboratory Tests

Safety laboratory data will include clinical chemistry, hematology, and urinalysis. 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.

Summaries in the form of shift tables for changes in CTCAE grades in key laboratory parameters showing the number and percentage of patients during the course of the study will be also be presented.

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. Graphs of key parametric clinical laboratory tests will be presented by treatment arm (mean ± standard deviation).
All laboratory values will be presented using conventional units.

15.5 ECOG

ECOG will be obtained at screening visit for all patients. Patients will be graded according to the ECOG Performance Status scale and criteria as described in Table 1. The ECOG will be summarized by grade and treatment arm using qualitative summary statistics and presented in listings.

Table 1: ECOG Performance Status

ECOG Scale	
0	Fully active, able to carry on all pre-disease performance without restriction
1	Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature (e.g., office work or light housework)
2	Ambulatory and capable of all self-care but unable to carry out any work activities; up and about more than 50% of waking hours
3	Capable of only limited self-care, confined to bed or chair more than 50% of waking hours
4	Completely disabled; cannot perform any self-care; totally confined to bed or chair

15.6 ECG

ECGs will be obtained at screening, Day 1 and 2 in Cycle 1, EOT, and at 30-day safety follow-up visit for all patients. The ECG will include heart rate, QRS, QT, and PR intervals, and will be presented in listings.

16. Interim Analyses

The study design is group sequential with 1 interim analysis (after 50 approximately patients in each treatment arm have completed at least 1 cycle in each of the treatment arms docetaxel + plinabulin [40 mg] versus docetaxel + pegfilgrastim, with matching placebos) and 1 final analysis at the completion of the study. These results assume that 2 sequential tests are made using the O'Brien-Fleming spending function to determine the test boundaries and the details of critical values and nominal alpha is defined in Table 2.

If non-inferiority is determined from the statistical testing, then also the hypothesis of superiority will be tested, and if it is concluded that the plinabulin treatment is superior to the pegfilgrastim treatment, with respect to DSN, then the study will be stopped. Since the design allows for stopping for inferiority, it might be decided on that occasion that the study will be stopped. Since this is a hierarchical testing procedure no penalty with respect to overall significance will be paid.

The statistical testing will be performed and will be reviewed by an independent Data Safety Monitoring Board (DSMB) at the interim analysis.

Table 2: Sequential testing parameters

Look #	# Patients	Upper Boundary	Nominal Alpha	Total Alpha	Total Power
1	100	2.50923	0.00605	0.00605	0.56076
2	150	1.99284	0.01895	0.025	0.90

The critical value of 2.50923 was used for calculating the 2-sided 95% confidence intervals obtained from the negative binomial regression model.

17. Changes from Protocol-Stated Analyses

The changes from the protocol-stated analyses are:

Protocol Amendment 6	**SAP 2.0**
NBR model will be implemented by using GENMOD procedure in SAS 9.4	NBR model will be implemented by using NBREG, GLM, and TABSTAT procedures in Stata v11.0. Through our research into the NBR models, we discovered that there may be up to more than 25 different implementations of the NBR model (Hilbe 2011). Joseph M. Hilbe whom we have collaborated with over the years was one of the greatest authorities on the NBR model and developed with others very many algorithms in Stata. SAS is only capable of just a small number of these models. Stata has more options within its software to adjust for dispersion, zero-inflation, and other aspects of the data distribution. One option in Stata is the dispersion function which allows for parameterization of dispersion. We have found that when the estimated alpha is used as the dispersion parameter, the variance is reduced notably. This makes for the most efficient NBR models and cannot be done in SAS.

18. References

Alexopoulous K, Kouroussis C, Androulakis N, et al. Docetaxel and granulocyte colonystimulating factor in patients with advanced non-small-cell lung cancer previously treated with platinum-based chemotherapy: a multicenter Phase II trial. Cancer Chemother Pharmacol 1999, 43: 257-262

Aarts M, Peters F, Mandigers C, et al. Primary granulocyte colony-stimulating factor prophylaxis during the first two cycles only or throughout all chemotherapy cycles in patients with breast cancer at risk for febrile neutropenia. J Clin Oncol. 2013 Dec 1;31(34); 4290-4296.

Blackwell K, Semiglazov V, Krasnozhon D, et al. Comparison of EP2006, a filgrastim biosimilar, to the reference: a Phase III, randomized, double-blind clinical study in the prevention of severe neutropenia in patients with breast cancer receiving myelosuppressive chemotherapy. Annals of Oncology 26: 1948–1953, 2015.

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

Deming WE. (1943). Statistical adjustment of data. Wiley, NY (Dover Publications edition, 1985).

Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A, et al. EORTC QLQ-C30 Scoring Manual. 3rd ed. Brussels: EORTC; 2001.

Giesinger JM, Wintner LM, Zabernigg A, Gamper EM, Oberguggenberger AS, Sztankay MJ, et al. Assessing quality of life on the day of chemotherapy administration underestimates patients' true symptom burden. BMC Cancer. 2014 Oct 10; 14:758.

Green MD, Koelbl H, Baselga J, Galid A, Guillem V, Gascon P, Siena S, Lalisang RS, Samonigg H, Clemens MR, Zani1V, Liang BC, Renwick J, Piccart MJ.Hanna N, Shepherd F, Fossella F, et al. Randomized phase III trial of pemetrexed versus docetaxel in patients with nonsmall-cell lung cancer previously treated with chemotherapy. J Clin Oncol. 2004 May 1;22(9):1589-97.

Hinz A, Einenkel J, Briest S, Stolzenburg J-U, Papsdorf K, Singer S. Is it useful to calculate sum scores of the quality of life questionnaire EORTC QLQ-C30 2012;21(5):677-683.

Holmes FA, O'Shaughnessy JA, Vukelja S, et al. Blinded, randomized, multicenter study to evaluate single administration pegfilgrastim once per cycle versus daily filgrastim as an adjunct to chemotherapy in high-risk stage II or stage III-IV breast cancer. J Clin. Oncol. 2002; 20:727731

Kaufman PA, Paroly W, Rinaldi D, et al. (2004). Randomized, double-blind, phase 2 study evaluating the safety of same vs next day administration of pegfilgrastim with docetaxel, doxorubicin and cyclophosphamide (TAC) in women with breast cancer. Breast Canc Res Treat; 88: S59, Abstract 1054

Lan KKG. and DeMets DL. 1983. 'Discrete sequential boundaries for clinical trials.' Biometrika, 70, pages 659-663.

Mehrotra DV, Liu F, Permutt T. Missing data in clinical trials: control‐based mean imputation and sensitivity analysis. Pharmaceutical Statistics. 2017; 16:378‐392.

Meza L, Baselga J, Holmes FA. et al, Incidence of febrile neutropenia (FN) is directly related to duration of severe neutropenia (DSN) after myelosuppressive chemotherapy. Proc Am Soc Clin Oncol. 2002; 21:255b (Abstr 2840).

Moore DC, Pellegrino AE. Pegfilgrastim-Induced Bone Pain: A Review on Incidence, Risk Factors, and Evidence-Based Management. Ann Pharmacother. 2017 Sep;51(9):797-803.

Nabholtz JM, Riva A., Taxane/Anthracycline Combinations: Setting A New Standard in Breast Cancer Oncologist. 2001; 6 Suppl 3:5-12

O'Brien PC and Fleming TR. 1979. 'A multiple testing procedure for clinical trials.' Biometrics, 35, pages 549-556.

van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health 2012;15(5):708-715.

Vogel C, Wojtukiewicz M, Carroll R, et al. First and subsequent cycle use of pegfilgrastim prevents febrile neutropenia in patients with breast cancer: a multicenter, double-blind, placebocontrolled phase III study. J Clin Oncol. 2005 Feb 20;23(6):1178-1184

Yoh K, Kasahara K, Yamada K, et al. A randomized, double-blind, phase II study of ramucirumab plus docetaxel vs placebo plus docetaxel in Japanese patients with stage IV nonsmall cell lung cancer after disease progression on platinum-based therapy. Lung Cancer. 2016 Sep; 99:186-93.

Zarixo® (filgrastim) FDA Oncologic Drugs Advisory Committee Meeting (FDA Advisory Committee Briefing Document). Jan 7, 2015.

19. Revision History

Version	Date
Draft SAP	20 July 2018
Version 1.0	26 December 2018
Version 2.0	20 May 2019

20. Summary Change of Statistical Analysis Plan

Section in SAP 2.0	Changes
List of abbreviations	Remove some unnecessary abbreviations and add some more abbreviations
2.1	Add the definition of DSN based on protocol amendment 6.
3.2	Make one correction: The Wilcoxon Rank Sum test for AUC will be applied by using the NPAR1WAY procedure.
3.3	Make one correction: Barnard's test will be applied by using the FREQ procedure.
15.1	Make a change to the definition of TEAE.
Through the whole document	Cosmetic minor changes