DISCLOSURE

REDACTED STATISTICAL ANALYSIS PLAN

ABI-007-NSCL-003

A PHASE III, RANDOMIZED, OPEN-LABEL, CROSSOVER, MULTI-CENTER, SAFETY AND EFFICACY STUDY TO EVALUATE NAB-PACLITAXEL (ABRAXANE®) AS MAINTENANCE TREATMENT AFTER INDUCTION WITH NAB-PACLITAXEL PLUS CARBOPLATIN IN SUBJECTS WITH SQUAMOUS CELL NON-SMALL CELL LUNG CANCER (NSCLC)

The information contained in the attached report is the property of Celgene and should not be shared or used for any purpose other than that for which it was provided.

Celgene is committed to providing information about its clinical trials to researchers and patients with the goal of furthering science and enhancing healthcare worldwide. Laws and regulations require, however, that Celgene protects patient privacy. The company may further have legal or contractual obligations not to disclose commercial or technical information provided by or related to certain partner companies or vendors.

The attached report is presented in its original format, but certain information has been redacted in order to comply with the aforementioned obligations or to protect Celgene's confidential commercial information. The redactions are based on the following principles:

- Redacted information has been replaced by grey space, maintaining original spacing and pagination.
- Any information that might allow the identification of individuals has been redacted for anonymization.
- Attachments to this report that contain confidential information are not made available. Such attachments include those that contain identifiable patient information, such as subject listings, narratives, and profiles. They also may contain confidential commercial information such as methodologies, and hypothesis generating and exploratory analyses.
- Cross-references to these attachments (such as links to subject listings in Section 16.2) are not redacted from the body of the report. However, the hyperlinks in the electronic document are no longer functional.
- Information about Celgene vendors and their services are redacted because many contracts prohibit disclosure of that information. Further, laws and regulations prevent us from disclosing certain information about our vendors or their services because it is protected by copyright.

Information about Celgene's redaction policies and the availability of additional data from this report may be found at http://www.celgene.com/research-development/clinical-trials/clinical-trials-data-sharing/.

STATISTICAL ANALYSIS PLAN

A PHASE III, RANDOMIZED, OPEN-LABEL, MULTI-CENTER, SAFETY AND EFFICACY STUDY TO EVALUATE *NAB*-PACLITAXEL (ABRAXANE®) AS MAINTENANCE TREATMENT AFTER INDUCTION WITH *NAB*-PACLITAXEL PLUS CARBOPLATIN IN SUBJECTS WITH SQUAMOUS CELL NON-SMALL CELL LUNG CANCER (NSCLC)

STUDY DRUG: *nab*-Paclitaxel (Abraxane®)

PROTOCOL NUMBER: **ABI-007-NSCL-003**

DATE FINAL: 19MAY-2017

Prepared by:

on behalf of

Celgene Corporation

86 Morris Avenue

Summit, NJ 07901

TABLE OF CONTENTS

LIST OF TABLES

SIGNATURE PAGE

STATISTICAL ANALYSIS PLAN (SAP) AND SAP AMENDMENT APPROVAL SIGNATURE PAGE	
SAP TITLE	ABI-007-NSCL-003 Statistical Analysis Plan
SAP VERSION, DATE	Draft 6, 28Feb2017
SAP AUTHOR	Signature and Date
PROTOCOL TITLE	A PHASE III, RANDOMIZED, OPEN-LABEL, MULTI-CENTER, SAFETY AND EFFICACY STUDY TO EVALUATE *NAB*-PACLITAXEL (ABRAXANE®) AS MAINTENANCE TREATMENT AFTER INDUCTION WITH *NAB*-PACLITAXEL PLUS CARBOPLATIN IN SUBJECTS WITH SQUAMOUS CELL, NON-SMALL CELL LUNG CANCER (NSCLC)
INVESTIGATIONAL PRODUCT	*nab*-PACLITAXEL (ABRAXANE®)
PROTOCOL NUMBER	ABI-007-NSCL-003
PROTOCOL VERSION, DATE	Amendment 3, 08Feb2017
SIGNATURE STATEMENT	By my signature, I indicate I have reviewed this SAP and find its contents to be acceptable.
Statistical Therapeutic Area Head	
Signature	
Printed Name	
Lead Clinical Resear	
Signature	
Printed Name	
Lead Product Safety Physician	

Signature	
Printed Name	

Signature	*{See appended electronic signature page}*		
Printed Name		Date	

1. LIST OF ABBREVIATIONS

Table 1: Abbreviations and Specialist Terms

Abbreviation	Meaning
ADT	Analysis date
AE	Adverse event
AESI	Adverse event of special interest
ALT	Alanine aminotransferase (SGPT)
ANC	Absolute neutrophil count
ANOVA	Analysis of variance
AST	Aspartate aminotransferase (SGOT)
AUC	Area under the curve
BMI	Body mass index
BSA	Body surface area
BSC	Best supportive care
CI	Confidence interval
CMH	Cochran-Mantel-Haenszel
CR	Complete response
CRF	Case report form
CT	Computerized tomography
CTCAE	Common Terminology Criteria for Adverse Events
DCR	Disease control rate
DMC	Data Monitoring Committee
ECOG	Eastern Cooperative Oncology Group
EOS	End of study
EQ-5D-5L	EuroQol 5-dimensional quality of life instrument
FNC	Fine cell cytology
Hgb	Hemoglobin
HR	Hazard ratio

Abbreviation	Meaning
IP	Investigational product
ITT	Intent-to-treat
IV	Intravenous
IVRS	Interactive voice response system
LCSS	Lung cancer symptom scale
MedDRA	Medical dictionary for regulatory activities
MRI	Magnetic resonance imaging
NA	Not applicable
NSCLC	Non-small-cell lung cancer
ORR	Overall response rate
OS	Overall survival
PD	Progressive disease
PFS	Progression-free survival
PN	Peripheral neuropathy
PP	Per protocol
PR	Partial response
PT	Preferred term
RE	Response evaluable
RECIST	Response evaluation criteria in solid tumors
SAE	Serious adverse event
SAP	Statistical analysis plan
SD	Stable disease
SE	Standard error
SGOT	Serum glutamic oxaloacetic transaminase (AST)
SGPT	Serum glutamic pyruvic transaminase (ALT)
SMQ	Standardized MedDRA query
SOC	System organ class
STDEV	Standard deviation

Abbreviation	Meaning
TEAE	Treatment emergent adverse event
TTP	Time to progression
ULN	Upper limit of normal
UE	Unevaluable
US	United States
VAS	Visual analog scale
WBC	White blood cell
WHO	World Health Organization

2. INTRODUCTION

This statistical analysis plan (SAP) describes the analyses and data presentations for Celgene's protocol ABI-007-NSCL-003 "A PHASE III, RANDOMIZED, OPEN-LABEL, MULTI-CENTER, SAFETY AND EFFICACY STUDY TO EVALUATE *NAB*-PACLITAXEL (ABRAXANE®) AS MAINTENANCE TREATMENT AFTER INDUCTION WITH *NAB*-PACLITAXEL PLUS CARBOPLATIN IN SUBJECTS WITH SQUAMOUS CELL NON-SMALL CELL LUNG CANCER (NSCLC)" which was issued on 27SEP2013 and amended on 28OCT2014, 07MAY2015 and 13MAR2017. It contains definitions of analysis populations, derived variables and statistical methods for the analysis of efficacy and safety.

These analyses include all the dry runs described in the Scope of Work, the non-inferential and non-binding interim analyses described in the protocol, and the final analysis for the clinical study report. The interim analyses will be performed by a different vendor. Throughout this SAP, for the Induction Part, the treatment arm will be referred to as the *nab*-Paclitaxel/carboplatin treatment, while in the Maintenance Part, the treatment arms will be referred to as the *nab*-Paclitaxel+BSC (referring to *nab*-paclitaxel plus best supportive care) and BSC (referring to best supportive care) arms. The purpose of the SAP is to ensure the credibility of the study findings by pre-specifying the statistical approaches to the analysis of study data prior to database lock and any data analysis for the first inferential interim analysis. This SAP is developed after the finalization of the protocol and will be finalized and signed 4 weeks prior to the clinical database lock for the first inferential interim analysis. All statistical analyses detailed in this SAP will be conducted using SAS® Version 9.2 or higher. The clinical cutoff date will be determined based on the occurrence of approximately 91 progression-free survival (PFS) events for the first non-binding interim analysis and approximately 136 PFS events for the final PFS analysis and interim overall survival (OS) analysis. The clinical cutoff for the final OS analysis will occur at 147 deaths, which is approximately 1.5 years after the final PFS analysis.

3. STUDY OBJECTIVES

3.1. Primary Objective

The primary objective of the study is to evaluate progression-free survival (PFS) with *nab*-paclitaxel as Maintenance treatment after response or stable disease (SD) with *nab*-paclitaxel plus carboplatin in subjects with squamous cell non small-cell lung cancer (NSCLC).

3.2. Secondary Objectives

The secondary objectives of the study are:

- To evaluate the safety and tolerability of *nab*-paclitaxel as Maintenance treatment after response or SD with *nab*-paclitaxel plus carboplatin in subjects with squamous cell NSCLC.
- To further assess the efficacy with *nab*-paclitaxel as Maintenance treatment after response or SD with *nab*-paclitaxel plus carboplatin in subjects with squamous cell NSCLC, as measured by secondary efficacy endpoints.

3.3. Exploratory Objectives

The exploratory objectives of the study are:

- To determine baseline tumor characteristics which predict response and resistance to *nab*-paclitaxel/carboplatin during the Induction Part of the study.
- To determine what changes in peripheral tumor characteristics during treatment are associated with acquisition of resistance in the subjects who had initial clinical benefit.
- To assess healthcare resource utilization during the Maintenance Part of the study.
- To assess the Lung Cancer Symptom Scale (LCSS) and EuroQol Group 5-Dimension Self-Report Questionnaire score (EQ-5D-5L).

4. INVESTIGATIONAL PLAN

4.1. Overall Study Design and Plan

This is a Phase 3, randomized, open-label, multi-center study of *nab*-paclitaxel plus best supportive care (BSC) or BSC alone as Maintenance treatment after response or SD with *nab*-paclitaxel plus carboplatin as Induction in subjects with squamous cell NSCLC. Approximately 450 subjects with Stage IIIB/IV squamous cell NSCLC eligible to receive standard treatment of *nab*-paclitaxel plus carboplatin for 4 cycles will be enrolled in the Induction Part of the study. If after the 4 cycles, the subject has a radiologically assessed complete response (CR), partial response (PR), or SD without clinical progression, they will be randomized (2:1) in the Maintenance Part of the study to receive *nab*-paclitaxel plus BSC or BSC alone. BSC is defined as the best palliative care per investigator (including but not limited to: antibiotics, analgesics, antiemetics, thoracentesis, pleurodesis, blood transfusions, nutritional support, and/or focal external-beam radiation for control of pain, cough, dyspnea, or hemoptysis) excluding antineoplastic agents. Approximately 216 subjects will be evaluable for the primary endpoint of PFS in Maintenance.

The study schematic is presented in Section 11.1.

4.2. Study Endpoints

4.2.1. Primary Endpoint

The primary endpoint is PFS in months from randomization into the Maintenance Part of the study to the start of disease progression based on investigator assessment using RECIST 1.1 guidelines or death.

4.2.2. Secondary Efficacy Endpoints

4.2.2.1. Secondary Efficacy Endpoints

Secondary efficacy endpoints specified in the protocol include the following:

- Overall survival (OS) from randomization into the Maintenance Part of the study
- Overall response rate (ORR) during the Induction and Maintenance Parts of the study based on investigator assessment according to RECIST 1.1 guidelines

4.2.2.2. Other Secondary Efficacy Endpoints

- PFS from Day 1 Cycle 1
- OS from Day 1 Cycle 1
- ORR in Maintenance Part beyond response in Induction Part
- Disease control rate (DCR) during the Induction Part and over the entire study

- Time to response during the Induction Part and over the entire study, with and without the requirement of confirmation of response
- Duration of response over the entire study, with and without the requirement of confirmation of response

4.2.3. Secondary Safety Endpoints

Safety parameters such as AE and SAEs.

4.2.4. Exploratory Endpoint(s)

- Healthcare resource utilization during the Maintenance Part of the study using a questionnaire including hospitalizations, emergency room visits, doctor or nurse visits, procedures, and/or additional medication.
- Change in the LCSS and EQ-5D-5L.

The following exploratory endpoints will be described in a separate SAP:

- The correlation between pretreatment tumor characteristics and response and resistance to the study treatment determined using next-generation sequencing methods, immunohistochemistry, or other analysis methods.
- The association between the changes in tumor characteristics and the acquisition of resistance to therapy from plasma samples taken at treatment failure during Maintenance.
- The association between DNA nucleotide polymorphisms in subject's germ line DNA and association with treatment efficacy and toxicity.

4.3. Stratification, Randomization, and Blinding

4.3.1. Induction Part

The Induction Part of the study is a one-arm, open label run-in period of *nab*-paclitaxel plus carboplatin that runs for four cycles (approximately 21 days per cycle). The main purpose of the Induction Part is to identify those subjects who are eligible for randomization in the Maintenance Part of the study. Approximately 450 subjects eligible for standard treatment with *nab*-paclitaxel plus carboplatin for 4 cycles will be enrolled, provided if all inclusion/exclusion criteria are met within a 28-day screening period prior to Cycle 1 Day 1.

Induction treatment will commence on Cycle 1 Day 1:

- *nab*-Paclitaxel 100 mg/m^2 IV infusion over 30 minutes on Days 1, 8, and 15 of each 21-day cycle
- Carboplatin AUC = 6 mg*min/mL IV on Day 1 of each 21-day cycle after completion of *nab*-paclitaxel infusion

If the subject has radiological or clinical PD, they will be discontinued from the study and their death dates will be provided if possible.

4.3.2. Randomization and Maintenance Part

Once 4 cycles have been completed for Induction, if the subject has a radiologically assessed CR, PR, or SD without clinical progression, he/she will be evaluated for eligibility for the Maintenance Part. Subjects (approximately 216) who satisfy all inclusion/exclusion criteria will be randomized 2:1 to receive either:

- *nab*-Paclitaxel 100 mg/m^2 IV infusion over 30 minutes on Days 1 and 8 of each 21-day cycle plus BSC until disease progression (for guidelines on starting dose for Maintenance Part of the study if the subject had dose reduced during the Induction Part of the study see Section 8.3 of protocol ABI-007-NSCL-003 dated DD Feb 2017)

OR

- BSC until disease progression

Maintenance therapy should start at the time of randomization. If this is not possible, a maximum of 7 days will be allowed from the date of randomization to the start date of Maintenance therapy. Subjects must start Maintenance therapy no earlier than 21 days and no later than 35 days from Day 1 of the fourth cycle of Induction therapy.

Randomization will be stratified by disease stage (IIIB vs IV), response in Induction (CR/PR vs SD), and Eastern Oncology Cooperative Group (ECOG) performance status (0 vs 1).

In cases where there is a stratification error at randomization, then the actual stratum from the CRF will be used for the final analysis.

A pooling algorithm will be employed to eliminate small stratums. Please refer to *Pooling Algorithm for Eliminating Small Stratums* in Section 7.3.11.2.

4.3.3. Follow-up Period

All subjects who discontinue from the Maintenance Part of the study for any reason other than withdrawal of consent, lost to follow-up, or death, will enter the Follow-up period that will have a visit 28 days after progression or discontinuation. Those subjects entering the follow up period without documented progression will continue to have CT scans in accordance with standard of care until documented progression of disease. Additionally, subjects will be followed for OS by phone approximately every 90 days for up to 1.5 year after final analysis of approximately 136 PFS events in which a total of 147 deaths is expected to be observed for the final OS analysis.

4.4. Sample Size Determination

Sample sizes for this part of the study are based on the pragmatic requirements for achieving the necessary PFS events in the Maintenance part of the study.

The Maintenance part of the study is designed to detect a hazard ratio (HR) of 0.60 for PFS improvement with the *nab*-paclitaxel plus BSC regimen over the BSC alone regimen. There is an 80% power to detect a hazard ratio of 0.60 using a two-sided test conducted at the 5% level of significance. This hazard ratio of 0.60 assumes an underlying exponential distribution for both treatment groups with a median time to PFS of 2 months for the BSC alone group and a median time to PFS of 3.33 months for the *nab*-paclitaxel plus BSC group. Hence, the final analysis will

be conducted after 136 PFS events (ie, events of disease progression or deaths from any cause) have occurred.

The final analysis of the primary endpoint will be performed when approximately 136 PFS events have been observed. A total of approximately 216 subjects will need to be randomized assuming an approximate 36 months accrual period and an approximate 2 months follow-up and an approximate 9% per month dropout rate.

The required number of PFS events and power calculation were obtained from

Approximately up to 450 subjects will be enrolled in the Induction part of the study. This sample size would provide approximately 216 subjects eligible for randomization in the Maintenance part.

In the *nab*-Paclitaxel Phase 3 NSCLC development study (Protocol CA031), 42% (13/31) and 63% (125/198) of the squamous subjects from the North American and ex-North American sites, respectively, were progression free after 4 cycles of treatment and remained in the study in Cycle 5 and beyond. It is assumed that approximately 50% (225/450) of the subjects in the Induction part of this study will meet the radiologic and clinical progression free criteria at the end of Cycle 4. Furthermore, it is expected that approximately 4% of these 225 subjects will not meet the other inclusion/exclusion criteria for the Maintenance part, yielding a total of approximately 216 subjects entering the Maintenance part of the study.

The actual number of subjects enrolling into the Induction part, and then being randomized in the Maintenance part of the study will be monitored on an ongoing basis and adjusted accordingly to ensure that the targeted number of required PFS events for the Maintenance part is attained. Enrollment of the Induction part, therefore, could be discontinued early or extended beyond the planned number of 450 subjects.

5. GENERAL STATISTICAL CONSIDERATIONS

Programming, statistical analysis, and reporting will be conducted according to standard operating procedures.

5.1. Reporting Conventions

The following reporting conventions apply generally to tables, listings, and figures:

- The study part will be included as a column in the listings where applicable.
- Data from all study centers will be combined for analysis;
- All stratified efficacy analyses will use the stratification factors for randomization (disease stage (IIIB vs IV), response in Induction (CR/PR vs SD), and ECOG performance status at randomization (0 vs 1));
- All statistical tests of the treatment effect will be interpreted based on the stopping boundaries at each pre-specified interim analyses; P-values will be rounded to 4 decimal places. P-values that round to 0.0000 will be presented as '<0.0001' and p-values that round to 1.0000 will be presented as '>0.9999';
- Confidence intervals (CIs) will be presented as 2-sided 95% CIs unless specified differently for a given analysis;
- Summary statistics will consist of the number and percentage of subjects (or cycles, if appropriate) in each category for discrete variables, and the sample size, mean, median, standard deviation, minimum, and maximum for continuous variables;
- All mean and median values will be formatted to one more decimal place than the measured value. Standard deviation values will be formatted to two more decimal places than the measured value. Minimum and maximum will be formatted to the same decimal place as the measured value.
- All percentages will be rounded to one decimal place. The number and percentage of responses will be presented in the form XX (XX.X%), where the percentage is in the parentheses;
- All listings will be sorted for presentation in order of randomized treatment arm, study center, subject, and date of procedure or event. Subjects who are not randomized will be listed as a non-randomized group before randomized subjects;
- All analysis and summary tables will have the analysis population sample size (i.e., number of subjects)

6. PROTOCOL DEVIATIONS/VIOLATIONS

The protocol deviations/violations will be identified and assessed by clinical research physician or designee following company standard operational procedure. Protocol deviations and violations will be reviewed before database lock to determine the per protocol population (defined in Section 7.3.1.3 for the Maintenance Part of the study). Events that could trigger exclusion from the per protocol population include inclusion/exclusion criteria violations, failure to take any study drug as assigned, randomization errors, and prohibited concomitant medications and procedures.

The protocol deviations/violations will be summarized separately by treatment arm for the ITT population. A by-subject listing of subjects with protocol violations in the ITT population will be provided.

7. STATISTICAL ANALYSIS

Statistical analyses are presented below for Induction and Maintenance Parts separately. Analyses such as demography are similar for the two parts of the study, but are repeated to make each analysis description self-contained.

7.1. Subject Disposition

The total number of subjects screened will be presented, and subjects with screen failure and reasons for screen failure will be summarized by frequency and percentage of total number of subjects screened.

The number of subjects in Induction ITT and Safety populations will be presented with percentages based on the number of subjects in the ITT population.

The number of subjects who are randomized to the Maintenance Part as well as the number who are not randomized to the Maintenance Part will be presented. Of those who are not randomized, the number of subjects who do not screen fail but never receive treatment, along with the number of treated subjects discontinue treatment during the Induction Part and reasons for treatment discontinuation will be summarized for all subjects who are deemed eligible for the study with the following categories:

- Death
- Adverse event
- Progressive disease
- Symptomatic deterioration
- Withdrawal by subject
- Lost to follow-up
- Study terminated by sponsor
- Protocol violation
- Other

Reasons for study discontinuation will be summarized for all subjects who sign informed consent and discontinue study during the Induction Part with the following categories:

- Screen failure
- Death
- Adverse event
- Withdrawal by subject
- Lost to follow up
- Protocol violation

- Other

The size of following analysis populations for the Maintenance Part will be summarized based on number of subjects randomized in the Maintenance Part:

- ITT Population
- Response Evaluable Population
- Safety Population
- PP Population

Subjects discontinuing the Maintenance treatment will be summarized as well as reasons for treatment discontinuation. Subjects who discontinue from the study during the Maintenance Part will be summarized as well as reasons for study discontinuation.

Finally, the number of subjects who discontinue Maintenance and enter into the follow up part of the study will be summarized by counts and percentage based on those who are randomized in the Maintenance Part.

Listings will be provided for discontinued subjects with reason for treatment and study discontinuation.

7.2. Induction Part

7.2.1. Study Population Definitions

7.2.1.1. Intent-to-Treat Population

The efficacy analysis will be performed on the ITT population, which includes all enrolled subjects who received at least 1 dose of study treatment regardless of whether the subject has any efficacy assessments collected.

7.2.1.2. Safety Population

The safety population will be the analysis population for all safety/ tolerability analyses. The safety population includes all subjects enrolled who received at least 1 dose of study treatment. Only subjects with clear documentation that no study treatment was administered will be excluded from the safety population.

7.2.2. Baseline values

For the Induction Part of the study, baseline is defined as the last non-missing value on or before the first dose of administration of *nab*-paclitaxel or carboplatin. If multiple values are present for the same date, the average of these values will be used as the baseline.

7.2.3. Demographics

Demographics will be summarized for the ITT population. Age (years) and weight (kg) at baseline will be summarized descriptively. Age category (<65 versus ≥65 years, <70 versus ≥70

years, and < 75 versus ≥ 75 years), sex, race, and ethnicity will be summarized by frequency counts.

Age will be calculated as follows: age = greatest integer ≤ [(date of informed consent – date of birth + 1) / 365.25].

7.2.4. Baseline Characteristics

Baseline characteristics, the latest collected on or before the first date of dosing, will be presented for the Induction Part of the study:

- ECOG performance status at baseline (0 or 1)
- Physician assessment of peripheral neuropathy (PN) at baseline
- Assessment of diabetes in medical history, which will support subgroup analyses of diabetic subjects

7.2.5. Lung Cancer History

The following items will be summarized for lung cancer history:

- Disease stage before administration of Induction therapy (IIIB or IV)
- Method of specimen collection (biopsy, surgical specimen, FNC, other)
- Confirmation of squamous cell histology and type (papillary, clear cell, basaloid, and other)
- Time from latest systemic anti-cancer therapy to first dose date in months
- Time from latest cancer therapy (last to occur of systemic anti-cancer therapy, radiation, or cancer surgery) to first dose date in months.

Subject listings will be provided for all of the above, as well as date of specimen collection.

7.2.6. Medical History

A summary of medical and surgical history will be presented by MedDRA system and organ class (SOC) and preferred term (PT). A similar summary will be generated for the currently active abnormalities only, by SOC and PT.

7.2.7. Prior and Concomitant Anti-Cancer Therapy

7.2.7.1. Prior Systemic Anti-Cancer Therapy

Prior systemic therapies (e.g. chemotherapy) will be summarized by therapeutic drug classes and generic drug names using the World Health Organization (WHO) drug dictionary. The duration, whether the number of cycles is known, the number of cycles (if known), and the best response will be presented in a listing.

7.2.7.2. Prior and Concomitant Radiation Therapy

The number and percentage of subjects who had any prior or concomitant radiation therapy will be presented. For subjects with prior and concomitant radiation therapy, the number and

percentage of subjects with each treatment site of radiation therapy will be presented. Duration, dose, fraction (where known), intent (adjuvant, curative, palliative, unknown) and setting (stand-alone, concurrent with other anti-cancer therapy, sequential to other anti-cancer therapy) will be summarized.

Prior radiation therapies will be listed.

7.2.7.3. Prior and Concomitant Non-small Cell Lung Cancer Surgeries

The number and percentage of subjects who had any prior or concomitant non-small cell lung cancer surgery will be presented by system organ class and preferred term. Prior and concomitant surgeries will be listed.

7.2.8. Prior and Concomitant Medications

Medications reported on the Prior and Concomitant medications CRF pages will be coded to therapeutic drug classes and generic drug names using the WHO drug dictionary version WHO-DDE Q3 2013. Medications initiated prior to the start of study treatment and continued after the start of study treatment will be counted as both prior and concomitant medications.

7.2.8.1. Prior Medications

Prior medications are defined as all medications that were started before Day 1 of treatment for the Induction Part. A summary showing the number and percentage of subjects who took prior medications will be presented by WHO drug dictionary therapeutic drug class and generic drug name. This summary will be presented for the ITT population.

7.2.8.2. Concomitant Medications

Concomitant medications for the Induction Part are defined as any non-study medications that were taken on/after the date of the first dose of study drug and within 28 days after the date of treatment discontinuation if the subject discontinues without being randomized, or taken on/after the date of the first dose of study drug and before the randomization date if the subject is randomized into the Maintenance Part.

Summaries showing the number and percentage of subjects who took concomitant medications will be presented by WHO therapeutic drug class and generic drug name will be presented for the ITT population.

7.2.9. Study Treatments and Extent of Exposure

All study treatment and extent of exposure summaries will be provided based on the safety population. Descriptive statistics will be provided for treatment duration, number of cycles, cumulative dose, dose intensity, and percentage of protocol dose.

7.2.9.1. Treatment and Cycle Start and End Dates

Induction treatment will commence on Day 1 Cycle 1, and planned cycle lengths are 21 days. Day 1 of treatment for the Induction Part is defined as the first day of any study drug.

Cycle Start and End Dates of Induction Part

Cycle start dates in the Induction Part are the first day of any Day 1 carboplatin administration (as recorded on the eCRF). Cycle end dates in the Induction Part are defined as the day before Day 1 of the following cycle. The treatment end date defined below is the Cycle 4 end date.

Treatment End Date of Induction Part

Treatment end date for the Induction Part, and the end date of the last cycle in the Induction Part, will be calculated as follows:

- For subjects who are not randomized, the maximum of date of last *nab*-paclitaxel dose + 6 and date of last carboplatin dose + 20 will be used as the **treatment end date**.
- For subjects who are randomized:
    - If the last dose is Day 1 of Cycle 4, then the minimum of Day 1 plus 20 days and the day before randomization date will be used for the treatment end date for the Induction Part.
    - If the last dose is Day 8 of Cycle 4, then the minimum of Day 8 plus 13 days and the day before randomization date will be used for the treatment end date for the Induction Part.
    - If the last dose is Day 15 of Cycle 4, then the minimum of Day 15 plus 6 days and the day before randomization date will be used for the treatment end date for the Induction Part.
        - In the event that the last dose is administered on the same day as randomization, then the treatment end date will be the randomization date.

The cycle number for each date of interest, e.g., AE start date, will be calculated based on the cycle window set by the start and end dates of the study medication.

7.2.9.2. Dosing Period for Each Study Drug

Dosing periods will be defined for each study drug and will be used as the denominators for the calculation of dose intensity for each drug.

Conceptually, the dosing period for a study drug is the time period starting at the first dose of the study drug and ending a protocol- and subject-specific number of days after the last dose, depending on the dosing schedule for the treatment arm and the subject's status. Ending the dosing period on the date of last dose would lead to an overestimation of the dose intensity. To avoid such overestimation, additional days of dosing period beyond the date of the last dose allow for the calculated dose intensity to reflect the degree to which the subject's dosing schedule aligned with the protocol-specified dose intensity. A subject's dosing period never extends beyond the date of death, but it can extend beyond the treatment discontinuation date in some circumstances.

For subjects who participate in the Maintenance Part, *nab*-paclitaxel dosing periods in the Induction Part are equal to treatment duration for the Induction Part. See Sections 7.2.9.1 and 7.2.9.3. For subjects who participate in the Maintenance Part, carboplatin dosing period is the

minimum of the study day of the Cycle 4 Day 1 dose of carboplatin plus 20 and the day number of the day before randomization.

For subjects who are still on treatment in the Induction Part at the time of clinical cutoff and subjects who discontinue during the Induction Part prior to the clinical cutoff date, the specific calculations of the Induction Part dosing period for each study drug and arm are given in Table 2 below:

Table 2: Calculation of Induction Dosing Period for Each Study Drug Based on Cycle Day Number of Last Dose of the Study Drug

(For subjects who are still on treatment in the Induction Part at the time of clinical cutoff and subjects who discontinue during the Induction Part prior to the clinical cutoff date)

***nab*-Paclitaxel**	**Carboplatin**
Last dose record is D1, D8, or D15	Last dose record is D1
min{study day of death date (if it exists), study day of date of last dose of *nab*-paclitaxel + 6}	min{study day of death date (if it exists), study day of date of last dose of carboplatin +20}
where the last dose is the latest exposure record of the respective treatment regardless of the actual dose or whether the dose was administered.	

7.2.9.3. Treatment Duration

Treatment duration (in weeks) is defined as:

[(Treatment end date for the Induction Part) – (Day 1 of treatment for the Induction Part) + 1]/7.

Cycle duration is defined as the time period from Day 1 of each cycle to the day prior to Day 1 of subsequent cycle; for the last Induction cycle, the end date is the treatment end date for the Induction Part as defined in Section 7.2.9.1.

Descriptive statistics will be provided for treatment duration and total number of cycles for both carboplatin and *nab*-paclitaxel. Number of subjects will also be tabulated by total number of cycles received for both carboplatin and *nab*-paclitaxel.

7.2.9.4. Cumulative Dose

Cumulative dose is defined as the sum of the values entered on the dose assigned field on the exposure eCRF, taken across the Induction Part in mg/m^2 for *nab*-paclitaxel or mg*min/mL for carboplatin. Cumulative dose will be computed separately for *nab*-paclitaxel and carboplatin. Only doses that were actually administered will be included in the cumulative dose. Descriptive statistics will be presented for cumulative dose for the safety population.

7.2.9.5. Dose Intensity

Dose intensity for a study drug is defined as the cumulative dose divided by the dosing period in weeks for that study drug. Dose intensity will be calculated separately for *nab*-paclitaxel and carboplatin.

Dose intensities will be calculated as follows:

- Dose intensity for carboplatin (mg*min/mL/wk) = [cumulative dose for carboplatin in mg*min/mL]/[carboplatin dosing period in weeks];
- Dose intensity for *nab*-paclitaxel (mg/m^2/wk) = [cumulative dose for *nab*-paclitaxel in mg/m^2]/[*nab*-paclitaxel dosing period in weeks].

Dose intensity will be presented by drug for the Induction Part of the study for the safety population.

7.2.9.6. Exposure, Dose Reduction/Delay, and Doses Not Administered

7.2.9.6.1. Dose Reduction

A dose reduction is defined as a dose administered after the Cycle 1 Day 1 dose which is at a lower dose level than the dose the subject received at the previous dosing visit.

7.2.9.6.2. Dose Not Administered

A dose not administered is any scheduled dose that is not administrated.

7.2.9.6.3. Dose Delay

A dose is considered delayed if the dose was administered later than the expected date prescribed by the protocol in the Tables of Events, given the consideration that an administrative window of ± 2 days is permitted for all visits except C1D1.

Subjects are to be dosed with either *nab*-paclitaxel on Day 1, 8, and 15 and with carboplatin on D1 in each of the 4 cycles in the Induction Part.

7.2.9.6.3.1. Definitions of actual dose date, reference date, and expected dose date

Assuming all the dose and visit dates are available and entered in the database as described in the CRF completion guidelines, then the definitions below provide the precise method for the determination of a dose delay. If any of the dates required for the determination of a dose delay is missing, then the imputation rules in the next section (Section 7.2.9.6.3.2) will be applied to facilitate the calculations.

The **actual dose date** is defined as the date when at least one of the IPs was administered, or when all the planned IPs were not administered as entered by the site in the Exposure CRF page and the visits module.

- If more than one IP was to be administered at a dose visit, but the IPs were given at different dates, then the latest dose dates among the IPs will be taken as the actual dose

date for that visit. Similarly, if an IP is given more than one time for a visit, then the latest of the dose dates will be taken as the actual dose date for that visit.

- For the calculation of dose delay parameters, if none of the planned IPs was administered at a dose visit, then that dose visit will be considered for "dose not administered" calculations only, but not for "dose delay" calculations.

The **reference date** is defined as the previous visit at which the patient was to be dosed with any of the IPs. For example, the reference date for Day 8 is D1.

The expected dose date for the current visit is defined as 7+2 days after the reference date for nab-paclitaxel and carboplatin.

A **dose delay** is declared if the **actual dose visit date** is later than the **expected dose visit date.**

7.2.9.6.3.2. Imputation Rules for the Reference Date

For the determination of the **actual dose date,** the current dose date of the visit of concern should be utilized. If the dose date is not available, then the dose is considered "not administered". No further consideration should be given regarding dose delay for that visit.

For the determination of the **reference date,** use the previous dose date if available, or the previous visit date if the previous dose date is missing. If both are missing, then use the dose date from two visits prior to the current dose visit, or the corresponding visit date if the former is missing. If both are missing, then use the dose date from three visits prior to the current dose visit, or the corresponding visit date if the dose date is missing. This logic can be extended accordingly until a dose date or visit date can be identified for the calculation of the reference date for the current dose visit of concern.

It should be noted that there are two scenarios for which the dose date could be missing:

1. The patient was physically present at the clinic and some measurements were taken, but the IPs were not administered as prescribed by the protocol, then the dose date in the Exposure database will be missing while the visit date would have been entered and is available in the Visit module database.

2. If the patient was not physical present at the clinic, then both the dose date in the Exposure database and the visit date in the Visit module database will be missing.

Each dosing record of the IPs will be listed. If any overdose occurred and reported on the Investigational Drug Overdose CRF page, then it will be summarized (e.g., number of subjects with overdose on each IP, difference in planned duration of infusion and actual duration of infusion, difference in planned dose and actual dose, whether the overdose result in any adverse event), and will also be displayed in a listing.

Treatment exposure and dose reductions, delays, and doses not administered will be summarized as follows:

- Number of cycles and doses administered;
- Number and percentage of subjects with at least 1 dose reduction, number of dose reductions, and reasons (adverse event or other) for reduction, by cycle and overall;
- Number and percentage of subjects with at least one dose delay, number of dose delays, by cycle and overall.
- Number and percentage of subjects with at least 1 dose not administered, number of doses not administered, and reasons (adverse event or other) for dose not administered, by cycle and overall. If the reason for dose not administered is not given, it will be categorized as "Data Not Available" in summary tables.

7.2.9.7. Percentage of Protocol Dose

Percentage of protocol dose is the dose intensity divided by the protocol weekly dose, expressed as a percentage.

- Percentage of protocol dose = (dose intensity / protocol weekly dose) * 100%

The protocol weekly doses are as follows:

- *nab*-Paclitaxel: 100 mg/m^2/wk (total of 12 doses – 1 each for days 1, 8, and 15 for each cycle * 4 cycles / 12 weeks in duration)
- Carboplatin: 2 mg*min/mL/wk (6 mg*min/mL * total of 4 doses – 1 on day 1 of each of 4 cycles / 12 weeks in duration)

Percentage of protocol dose for carboplatin and *nab*-paclitaxel will be categorized into <20%, ≥20% to < 30%, ≥30% to < 40%, ≥40% to < 50%, ≥50% to < 60%, ≥60% to < 70%, ≥70% to < 80%, and ≥80 to <90% and ≥90%, and frequency counts will be provided for the safety population.

7.2.10. Efficacy Analysis

All efficacy evaluations will be conducted using the ITT population. Any presentation of response for the Induction Part of the study will be for descriptive or estimation purposes only.

The percent change from nadir during the Induction Part in sum of total length is required to assess target lesions for tumor progression. This is calculated as follows:

Percent change from nadir at a given visit= (total length – nadir in total length) / (nadir in total length) x 100%, where nadir is calculated using all measurements before the visit in question, including baseline.

Listings will be provided for all endpoints. To support RECIST 1.1 assessments, the following information will be presented in listings:

- Per subject visit:

- Date of assessment
- Number of target lesions
- Total length of target lesions
- Percent change from nadir of sum of total length (derived as above to review assessment of progressive disease)
- Percent change from baseline of sum of total length to review assessment of response
- Assessment of target lesions
- Assessment of non-target lesions
- Presence of symptomatic deterioration
- Overall tumor response assessment
- Best overall response

- Per lesion at each subject visit:
  - Location
  - Method of assessment
  - Tumor length (target lesions only)
- Overall per subject
  - The best response over the Induction Part

7.2.10.1. Tumor Response and Overall Response Rate

The primary purpose of the Induction Part is to identify subjects who achieve a radiologically assessed tumor response of SD, PR, or CR without clinical progression according to RECIST 1.1 criteria at the end of Cycle 4 and are therefore eligible for randomization into the Maintenance Part of the study.

The number and percentage of those with a tumor response of SD, PR, or CR will be presented, as well as the number and percentage with each of PD, SD, PR, CR, and unevaluable (UE). The denominator for the percentage will include all Induction Part ITT subjects, regardless of whether they had a tumor assessment. Subjects with no tumor assessments will be considered UE. This analysis will be presented for each protocol-specified CT scan (Week 6 for the end of Cycle 2, and Week 12 for the end of Cycle 4). If there is more than one CT scan within a relative day range, the best response will be counted. Relative study day windows are presented in Section 11.4.

The overall response rate is the percentage of subjects who achieved best response of confirmed PR or CR compared with baseline, where baseline is the last CT-scan obtained prior to or on Day 1 of treatment for the Induction Part. In this study, the best response of PR or CR has to be confirmed by repeated assessment no less than 28 days after the response criteria are met. The ORR along with 95% Clopper-Pearson confidence interval will be presented. The following rules will be observed for response rate in the Induction Part:

- For subjects who participate in the Maintenance Part, use best response among any tumor assessments on or before the day of randomization. The first radiographical evaluation performed within 91 days after randomization can be utilized as the confirmation scan, if necessary.
- For subjects discontinued prior to the Maintenance Part, use best response on or up until the patient receives a subsequent anti-cancer therapy
- All Induction Part ITT subjects will be included for this analysis.
- Confirmation of CR/PR will be programmed according to the standard guidance document created by the Biostatistics department.

7.2.10.2. Disease Control Rate

Disease Control rate is defined as the percent of subjects who have a radiologic complete response, partial response, or stable disease according to RECIST 1.1 guidelines, as determined by the investigator. Disease Control rate over the Induction Part along with associated Clopper-Pearson 95% CI will be presented for the Induction ITT population.

7.2.10.3. Time to Response

Time to response is defined as the time from the first dose of study drug to the first occurrence of confirmed response (CR or PR). Time to response will be summarized with descriptive statistics.

Only subjects with a CR or PR as a best overall confirmed response during Induction will be included in this analysis.

7.2.11. Safety Analyses

The purpose of this section is to describe the safety analyses for the Induction Part of the study. Adverse events and laboratories will be analyzed based on the safety population.

7.2.11.1. Adverse events

Adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) Version 17.0 or above. Appendix 11.5 gives the dates of application of new versions of MedDRA.

Adverse events will be analyzed in terms of treatment emergent adverse events (TEAEs), defined as any AEs that began or worsened in grade on or after Day 1 of treatment for the Induction Part and on or before the date of randomization for those who enter in the Maintenance Part, or through treatment discontinuation date plus 28 days for those who do not enter in the Maintenance Part. In addition, any SAE deemed by the investigator to be related to study drug is a TEAE, even if beyond the 28-day window.

For the summary of treatment related AEs, a treatment-related TEAE is defined as an adverse event which was considered to be related, reported as “suspected” in eCRF, to either component of combination drugs. Additionally, for related AEs reported as “suspected” in eCRF to each individual drug, *nab*-paclitaxel or Carboplatin will be summarized for each study drug. If a subject experiences multiple occurrences of the same AE with different relationship to study medication categories, the subject will be counted once, as a relationship category of treatment

related. AEs with a missing relationship will be presented in the summary table as a relationship category of "treatment-related".

The incidence of TEAEs/treatment related TEAEs will be summarized by MedDRA system organ class (SOC) and preferred term (PT) in a decreasing frequency order for SOC or PT. If a subject experiences multiple AEs under the same preferred term (system organ class), then the subject will be counted only once for that preferred term (system organ class). The intensity of AEs will be graded 1 to 5 according to the National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE) Version 4.0. For all other AEs not described in the CTCAE criteria, the intensity will be assessed by the investigator as mild (grade 1), moderate (grade 2), severe (grade 3), life-threatening (grade 4) or death (grade 5). If a subject experiences the same AE more than once with different toxicity grade, then the event with the highest grade will be tabulated in "by grade" tables. In addition, AEs with a missing intensity will be presented in the summary table as an intensity category of "Missing."

Tables summarizing the incidence of TEAEs will be generated for each of the following:

- Overall summary of TEAEs;
- TEAEs presented by MedDRA system organ class, preferred term, and worst CTCAE grade
- Treatment-related TEAEs by grade category (grade 1-2 vs. grade 3-4 vs. grade 5; overall and by cycle);
- Serious TEAE by worst CTCAE grade;
- Treatment-related serious TEAE by worst CTCAE grade;
- TEAEs with action of study drug withdrawn;
- Treatment-related TEAEs with action of study drug withdrawn;
- TEAEs with action of study drug dose reduced or interrupted;
- Treatment-related TEAEs with action of study drug dose reduced or interrupted;
- TEAEs with fatal outcome;
- Treatment-related TEAEs with fatal outcome;
- All deaths within 28 days of treatment discontinuation date with cause of death;
- Most frequent TEAEs (≥ 5%; sorted by Preferred Term, without SOC);
- TEAEs for the following baseline subgroups (provided the number of subjects are sufficient):
  - Age (<65 years versus 65 to ≤ 74 years versus ≥ 75 years)
  - Sex (male versus female)

For TEAEs that are summarized by cycle, the following rules will be used:

- TEAEs which occur on Day 1 Cycle 1 belong to Cycle 1.

- After cycle 1, TEAEs will be categorized by the "throw-back rule", that is, TEAEs that occur on Day 1 of a cycle will be allocated to the previous cycle.
- All TEAEs which occur after Day 1 of last cycle will be included in the last cycle. For subjects who are randomized, all TEAEs which occur after Day 1 of Cycle 4 and on or before the randomization date will be included in Cycle 4.

Data listings will be presented for all TEAEs. A listing for non-treatment-emergent AEs will also be provided.

7.2.11.2. Peripheral Neuropathy

Peripheral neuropathy events will be collected on the Physician Assessment of Peripheral Neuropathy eCRF and reported as AEs and will be included in analyses described in Section 7.2.11.1. A summary of PN grade by Induction cycle will be provided based on the eCRF data.

For subjects who were treated in the Induction Part only (i.e., subjects who discontinued treatment without being randomized), additional summaries of peripheral neuropathy events will be presented using both the AE data and the PN eCRF data. Adverse events will be considered PN events if they are TEAEs and are identified as peripheral neuropathy based on the Standardized MedDRA Queries (SMQs). The following will be summarized:

- Number and percentage of subjects who develop grade 3 or higher peripheral neuropathy;
- Time to first onset of grade 3 or higher peripheral neuropathy;
- Time to improvement of grade 3 or higher peripheral neuropathy by at least 1 grade; and
- Time to improvement of grade 3 or higher peripheral neuropathy to grade 0 or grade 1.

Due to the small number of events expected, time to first onset will be summarized with descriptive statistics for only those subjects who experienced peripheral neuropathy of the relevant grade. There will be no censoring.

Time to improvement of the grade ≥3 PN events listed above will be summarized for subjects who have a grade ≥3 PN event using the Kaplan-Meier methods. Subjects who do not achieve the relevant improvement by the time of the last evaluation of adverse events will be censored at the last time the subject is evaluated for adverse events.

7.2.11.3. Adverse Events of Special Interest

The following summaries will be provided for TEAEs included in the above-mentioned AEs of interest:

1. TEAEs by maximum CTCAE grade;
2. Serious TEAEs;
3. TEAEs with action of drug withdrawn;

4. TEAEs with action of dose reduced or interrupted;
5. TEAEs with fatal outcome;

TEAEs of special interest will be listed.

7.2.11.4. Clinical Laboratory Evaluations

Listings will be provided for the local laboratory data.

7.2.11.4.1. Hematology

All hematology data will be listed.

7.2.11.4.2. Clinical Chemistry

All clinical chemistry data will be listed.

7.2.11.5. Concurrent Radiation and Surgeries

The number of subjects having concomitant radiation or surgeries and procedures performed will be summarized.

Concomitant radiation or surgeries will be listed.

7.2.12. Quality of Life Analysis

The Lung Cancer Symptom Scale (Pol) and EQ-5D-5L questionnaires will be used to measure quality of life (QoL) for subjects in the trial. The LCSS is a 9 question assessment the subject completes using a visual analogue scale (VAS) to denote intensity of a symptom. The EQ-5D-5L comprises 5 questions on mobility, self-care, usual activities, pain/discomfort, anxiety/depression and a VAS for overall QoL. In addition, there is an overall index score that is derived from the 5 questions. These questionnaires will be completed at Day 1 of every cycle, and at Early Treatment Discontinuation/End of Cycle 4 during the Induction Part of the study. Baseline scores are defined as scores captured on Day 1 (first dose date). In the case of multiple assessments on Day 1, the later assessment will be used. Missing baseline scores will not be imputed.

For administrative purposes (e.g., planning of new studies, publication of data) the LCSS and EQ-5D measurements collected during the Induction Part will be summarized overall and by subgroup using descriptive statistics as needed when approximately 50, 100, 200, 400, and all subjects, respectively, have either discontinued during or completed the Induction Part of the study. Subgroups may include but are not limited to radiological response and baseline symptom status.

7.2.12.1. LCSS

The LCSS is designed to measure quality of life specifically affected by lung cancer and its symptoms. It evaluates burden due to six major symptoms:

- Appetite
- Fatigue
- Coughing

- Shortness of breath
- Blood in sputum
- Pain

In addition, it measures how bad a subject's symptoms are, how much lung cancer has affected normal activities, and global quality of life.

The LCSS consists of two scales: one to be filled out by the subject and one by the health care provider. For this study, only the portion filled out by the subject will be used. The subject will respond to each of the 9 items using marks on a 100 mm visual analog scale (VAS). For analysis and presentation purposes, the scores will be presented such that 0 mm corresponds to the worst possible health state and 100 mm corresponds to the best possible health state.

The average of the VAS score of all 9 items will be calculated for each subject and will be referred to as the LCSS total score. The symptom burden index is the average of the 6 symptom-specific items (the first 6 items). The average of the coughing, shortness of breath, and blood in sputum items will be referred to as the Respiratory Symptom Scale. The average of the appetite and fatigue items (the first 2 items) will be referred to as the overall constitutional score. The sum of the symptoms item (item 7), the normal activities item (item 8), and the global quality of life item (item 9) will be referred to as the general health scale.

Summary statistics of the individual LCSS items and summary scale scores, the change from baseline, and percent change from baseline will be summarized for the ITT population at every time point at which the instrument is given, and at end of Induction (i.e., last assessment during Induction or Cycle 5 Day 1, whichever is latest). In addition, the change from baseline to the LCSS score during the Induction Part, defined as the difference between the last reported LCSS score during the Induction Part, either the end of 4 cycles or early discontinuation visit or the last reported LCSS score before clinical cutoff for subjects who have not reached that point at an interim analysis, will be presented.

7.2.12.2. EQ-5D-5L

The EQ-5D-5L has been designed as an international, standardized, generic instrument for describing and valuing health-related quality of life. It contains only the domains common to generic health status measures, contains the minimum number of questions for each domain, was designed for ease of self-administration, and produces a single index for analysis.

Five dimensions are covered by the EQ-5D-5L:

- Mobility
- Self-care
- Usual Activities
- Pain/Discomfort
- Anxiety/Depression

The 5 dimensions are scored on a 5-point Likert scale. In addition, there is a question covering perceived change in health status over the previous 12 months measured on a 20 cm vertical

visual analog scale "thermometer" to summarize one's overall health status at administration on a 100-point scale.

Responses from the 5 dimensions are coded so that a '1' indicates no problem on that dimension, and '5' indicates the most serious problem. A profile is comprised of 5 digits consisting of 1s, 2s, 3s, 4s, and 5s. For instance, a profile of 11111 indicates no problem with any of the 5 dimensions, while a profile of 55555 indicates the most difficulty on all 5. Health state scores are based on the VAS with zero corresponding to the worst imaginable health state and 100 corresponding to the best imaginable health state.

A utility score will be calculated using the US Crosswalk Index Value Set obtained from the EuroQol website. The range of possible values for the utility score based on the US Crosswalk Index Value Set is -0.109 to 1.000. The overall utility score, change from baseline, and percent change from baseline in the overall utility score will be summarized for every time point at which the instrument is given for the ITT population.

Shifts from baseline to each time point at which the instrument is given will be presented for each dimension for the ITT population.

Summaries of the VAS results and the change and percent change from baseline will be presented for every time point at which the instrument is given.

7.3. Maintenance Part

7.3.1. Study Population Definitions

7.3.1.1. Intent-to-Treat Population

The primary efficacy analysis will be performed on the ITT population, which includes all randomized subjects regardless of whether the subject receives any study regimen or has any efficacy assessments collected. In all ITT analyses, subjects are analyzed according to their randomized treatment, regardless of what treatment they may receive at any time after randomization.

7.3.1.2. Response Evaluable Population

Response evaluable (RE) population includes all randomized subjects who meet eligibility criteria, take at least one dose of study drug or are on BSC and have at least one baseline and one post-randomization efficacy assessment. The response evaluable population will be used in a replication of the primary and major secondary efficacy analyses to assess the effect of subjects who were randomized but not treated or assessed on the efficacy analysis.

7.3.1.3. Per-Protocol Population

The per-protocol (PP) population, which will include all subjects randomized in the Maintenance part who receive at least one dose of the study therapy or on BSC arm and do not have any known major protocol deviations and fulfill the study enrollment criteria. The PP population will be used in a replication of the primary and major secondary efficacy analyses to assess the effect of major protocol deviations on efficacy. Protocol deviations and their handling for analysis purposes, are discussed in more detail in Section 6.

7.3.1.4. Safety Population

The safety population includes all randomized subjects. The safety population will be the basis of all safety analyses. Subjects who receive *nab*-paclitaxel in error will be analyzed by actual treatment received.

7.3.2. Demographics

Demographics will be summarized for the ITT population. Age (years) and weight (kg) at baseline will be summarized descriptively. Age category (<65 versus ≥ 65 years, <70 versus ≥70 years, and < 75 versus ≥ 75 years), sex, race, and ethnicity will be summarized by frequency counts.

Age will be calculated as follows: age = Integer ≤ [(Date of informed consent – Date of Birth + 1) / 365.25].

7.3.3. Baseline Characteristics

The following characteristics collected before the beginning of the Induction Part will also be presented for the subjects in the Maintenance Part of the study by treatment arm for the ITT population:

- Assessment of diabetes in medical history, which will support subgroup analyses of diabetic subjects

In addition, the following measures collected before the Maintenance Part of the study will be summarized. The last non-missing measure before randomization be considered baseline for these measures:

- ECOG performance status at end of Induction (0 or 1)
- Physician assessment of peripheral neuropathy (PN)
- The tumor response (CR, PR, and SD) achieved in the Induction Part

7.3.4. Lung Cancer History

The following items will be summarized by treatment arm for lung cancer history:

- Disease stage before administration of Induction therapy (IIIB or IV)
- The time from the first diagnosed date of lung cancer (specimen collection date if the diagnosed date is unknown) to Day 1 of treatment for the Induction Part in months, defined as (Day 1 of treatment for the Induction Part – the first diagnosed date (or specimen collection date if the diagnosed date is unknown) + 1)/30.4375
- Method of specimen collection (biopsy, surgical specimen, fine needle cytology (FNC), other)
- Confirmation of squamous cell histology and type (papillary, clear cell, basaloid, and other)

Subject listings will be provided for all of the above, as well as date of specimen collection.

7.3.5. Medical History

A summary of medical and surgical history will be presented by MedDRA system and organ class (SOC) and preferred term (PT). A similar summary will be generated for the currently active abnormalities only at randomization, by SOC and PT.

7.3.6. Prior and Concomitant Anti-Cancer Therapy

Subject listings described in the Induction Part of this SAP will include a column for treatment group, if applicable.

7.3.6.1. Prior Systemic Anti-Cancer Therapy

The number and percentage of subjects with any prior systemic anti-cancer therapy will be presented by treatment group for the ITT population.

7.3.6.2. Prior and Concomitant Radiation Therapy

The number and percentage of subjects who had any prior or concomitant radiation therapy will be presented by treatment group for the ITT population. For subjects with prior and concomitant radiation therapy, the number and percentage of subjects with each site of radiation therapy will be presented.

7.3.6.3. Prior and Concomitant Non-small Cell Lung Cancer Surgeries

The number and percentage of subjects with any prior or concomitant cancer surgery will be presented by treatment group, system organ class and preferred term for the ITT population. The time from the last prior non-small cell lung cancer surgery to date of first dose will be summarized.

7.3.7. Prior Medications

Prior medications are defined as in the Induction Part. A summary showing the number and percentage of subjects randomized to the Maintenance Part who took prior medications will be presented by WHO drug dictionary therapeutic drug class and generic drug name. This summary will be presented for the ITT population.

7.3.8. Concomitant Medications

Concomitant medications for the Maintenance Part are defined as non-study medications that were taken on/before the randomization and within 28 days of the treatment discontinuation date.

Summaries showing the number and percentage of subjects who took concomitant medications by WHO therapeutic drug class and generic drug name will be presented for the ITT population.

7.3.9. Concurrent Radiation and Surgeries

The number of subjects having concomitant surgeries and procedures performed will be summarized.

Concomitant surgeries will be listed.

7.3.10. Study Treatments and Extent of Exposure

All study treatment and extent of exposure summaries will be provided based on the safety population. Descriptive statistics will be provided for treatment duration, number of cycles, cumulative dose, dose intensity, and percentage of protocol dose by treatment arm.

7.3.10.1. Treatment and Cycle Start and End Dates

Maintenance treatment will commence on Day 1 Cycle 5, and planned cycle lengths are 21 days. Day 1 of treatment for the Maintenance Part is defined as the date when the Day 1 Cycle 5 nab-paclitaxel dose was administered for the nap-paclitaxel arm, and the first visit date of any Day1 Cycle 5 assessments for the BSC arm.

Cycle Start and End Dates of Maintenance Part

For subjects randomized to nab-paclitaxel + BSC, Day 1 of Cycle 5 is defined as the date when the Day 1 Cycle 5 nab-paclitaxel dose was administered and Day 1 of any cycle beyond Cycle 5 is the date of Day 1 nab-paclitaxel administration in a cycle. For subjects randomized to BSC alone, Day 1 of Cycle 5 is defined as the first visit date of any Day1 Cycle 5 assessments and Day 1 of any cycles beyond Cycle 5 is the first visit date of any Day 1 assessments for that cycle. Cycle end dates are the day before Day 1 of the following cycle. The treatment end date described in the section below is the end date for the last cycle.

Cycle duration is defined as the time period from Day 1 of each cycle to one day prior to the Day 1 of subsequent cycle (see details in Section 5.2.2).

Treatment End Date of Maintenance Part

Treatment end date for the Maintenance Part, is defined as the date before the next planned cycle.

The cycle number for each date of interest, e.g., AE start date, will be calculated based on the cycle window set by their start and end dates.

7.3.10.2. Dosing Period for *nab*-Paclitaxel

The Maintenance Part dosing period will be defined for *nab*-paclitaxel and will be used as the denominator for the calculation of dose intensity.

Conceptually, the dosing period for a study drug is the time period starting at the first dose of a study drug and ending at subject-specific number of days after the last dose, depending on the dosing schedule for the treatment arm and the subject's status. Ending the dosing period on the date of last dose would lead to an overestimation of the dose intensity as the exposure to the study drug extends beyond the dosing date. To avoid such overestimation, additional days of dosing period beyond the date of the last dose allow for the calculated dose intensity to reflect the degree to which the subject's dosing schedule aligned with the protocol-specified dose intensity. A subject's dosing period for a study drug never extends beyond the date of death, but it can extend beyond the treatment discontinuation date in some circumstances. The specific calculations of the Maintenance Part dosing period are given in Table 4 below:

Table 4: Calculation of Maintenance *nab*-Paclitaxel Dosing Period Based on Cycle Day of Last Dose

Last dose record is D1	**Last dose record is D8**
min{study day of death date (if it exists), study day of date of last dose of *nab*-paclitaxel + 9.5}	min{study day of death date (if it exists), study day of date of last dose of *nab*-paclitaxel + 13}
where the last dose is the latest exposure record of *nab*-paclitaxel regardless of the actual dose or whether the dose was administered.	

7.3.10.3. Treatment Duration

Treatment start date for the Maintenance Part is defined as the date when the Day 1 Cycle 5 nab-paclitaxel dose was administered for the nap-paclitaxel arm, and the first visit date of any Day1 Cycle 5 assessments for the BSC arm. For the BSC arm, treatment duration is an indicator for the duration the subject participated in the Maintenance period prior to discontinuation due to death, adverse event, progressive disease, symptomatic deterioration, withdrawal by subject, lost to follow up, protocol violation, or other reason, although no study regimen was administered.

Treatment duration (in weeks) is defined as:

[(Treatment end date for the Maintenance Part) – (Treatment start date for the Maintenance Part) + 1]/7.

Descriptive statistics will be provided for treatment duration and total number of cycles for the *nab*-paclitaxel+BSC arm, both for the Maintenance Part and for the entire study. Number of subjects will also be tabulated by total number of cycles received in the Maintenance Part and in the entire study.

7.3.10.4. Cumulative Dose

Cumulative dose is defined as the sum of the values entered on the dose assigned field on the exposure eCRF, taken across the study period in mg/m^2 for *nab*-paclitaxel. Only doses that were actually administered will be included in the sum.

As subjects who entered the Maintenance Part might have a different dose intensity profile than subjects who did not, descriptive statistics will be presented for cumulative dose over the Induction Part (by Maintenance treatment arm) and Maintenance Part for the *nab*-paclitaxel+BSC arm for the safety population (subjects treated) in the Maintenance Part. In addition, cumulative dose for carboplatin from the Induction Part, calculated as described in Section 7.2.9.4, will be summarized by Maintenance treatment arm for the same safety population in the Maintenance Part.

7.3.10.5. Dose Intensity

Dose intensity of *nab*-paclitaxel during the Maintenance Part or entire study is defined as the cumulative dose divided by the dosing period for the Maintenance Part or the entire study, respectively. Dose intensity of carboplatin will be calculated for the Induction Part only.

Dose intensities will be calculated as follows:

- Dose intensity for *nab*-paclitaxel = [cumulative dose for *nab*-paclitaxel in mg/m^2]/[dosing period in weeks].

Dose intensity for *nab*-paclitaxel will be presented for the safety subjects in *nab*-paclitaxel+BSC. In addition, dose intensity from the Induction Part, described in Section 7.2.9.5, will be summarized by Maintenance treatment arm including only Maintenance safety subjects.

7.3.10.6. Exposure, Dose Reduction/ Delay, and Doses Not Administered

7.3.10.6.1. Dose Reduction

A dose reduction is defined as a dose administered after the Cycle 5 Day 1 dose which is at a lower dose level than the dose the subject actually received at the previous dosing visit for the nab-paclitaxel + BSC arm.

7.3.10.6.2. Dose Not Administered

A dose not administered is any scheduled dose that is not administrated.

7.3.10.6.3. Dose Delay

A dose is considered delayed if the dose was administered later than the expected date prescribed by the protocol in the Tables of Events, given the consideration that an administrative window of ± 2 days is permitted for all visits except C5D1.

Subjects are to be dosed with nab-paclitaxel on Day 1 and 8 in the Maintenance Part.

7.3.10.6.3.1. Definitions of actual dose date, reference date, and expected dose date

Assuming all the dose and visit dates are available and entered in the database as described in the CRF completion guidelines, then the definitions below provide the precise method for the determination of a dose delay. If any of the dates required for the determination of a dose delay is missing, then the imputation rules in the next section (Section 7.2.9.6.3.2) will be applied to facilitate the calculations.

The **actual dose date** is defined as the date when nab-paclitaxel was administered, or when the planned dose was not administered as entered by the site in the Exposure CRF page and the visits module.

- If nab-paclitaxel is given more than one time for a visit, then the latest of the dose dates will be taken as the actual dose date for that visit.

- For the calculation of dose delay parameters, if nab-paclitaxel was not administered at a dose visit, then that dose visit will be considered for “dose not administered” calculations only, but not for “dose delay” calculations.

The **reference date** is defined as the previous visit at which the patient was to be dosed with nab-paclitaxel. For example, the reference date for Day 8 is D1 and the reference date for Day 1 of a cycle is D8 of the previous cycle.

The **expected dose date** for the current visit is defined as 7+2 days after Day 1 and 14+ 2 days after Day 8.

A **dose delay** is declared if the **actual dose visit date** is > the **expected dose visit date.**

7.3.10.6.3.2. Imputation Rules for the Reference Date

For the determination of the **actual dose date,** the dose date of the visit of concern should be utilized. If the dose date is not available, then the dose is considered "not administered". No further consideration should be given regarding dose delay for that visit.

For the determination of the **reference date,** use the previous dose date if available, or the previous visit date if the previous dose date is missing. If both are missing, then use the dose date from two visits prior to the current dose visit, or the corresponding visit date if the dose date is missing. If both are missing, then use the dose date from three visits prior to the current dose visit, or the corresponding visit date if the dose date is missing. This logic can be extended accordingly until a dose date or visit date can be identified for the calculation of the reference date for the current dose visit of concern.

It should be noted that there are two scenarios for which the dose date could be missing:

1. The patient was physically present at the clinic and some measurements were taken, but nab-paclitaxel was not administered as prescribed by the protocol, then the dose date in the Exposure database will be missing while the visit date would have been entered and is available in the Visit module database.

2. If the patient was not physical present at the clinic, then both the dose date in the Exposure database and the visit date in the Visit module database will be missing.

Each dosing record of nab-paclitaxel will be listed. If any overdose occurred and reported on the Investigational Drug Overdose CRF page, then it will be summarized (e.g., number of subjects with overdose on nab-paclitaxel, difference in planned duration and actual duration, difference in planned dose and actual dose, whether the overdose result in any adverse event), and will also be displayed in a listing.

7.3.10.7. Percentage of Protocol Dose

Percentage of protocol dose is the dose intensity divided by the protocol weekly dose, expressed as a percentage.

- Percentage of protocol dose = dose intensity / protocol weekly dose * 100%

For subjects who start Maintenance therapy with an assigned dose of 100 mg/m^2, the protocol weekly dose for *nab*-paclitaxel during the Maintenance Part is 66.7 mg/m^2/week (the subject will receive two 100 mg/m^2 doses of *nab*-paclitaxel on Days 1 and 8 of the 21-day cycle).

While it is possible that a starting dose for the Maintenance Part of 75 mg/m^2 may be given for subjects who had two dose reductions during Induction, the protocol weekly dose is 66.7 mg/m^2/week for all subjects randomized to *nab*-paclitaxel + BSC.

Percentage of protocol dose for *nab*-paclitaxel will be categorized into 10% intervals, and frequency counts will be provided for the safety population.

Treatment exposure and dose reductions, delays, and interruptions will be summarized for the *nab*-paclitaxel+BSC arm as follows:

- Number of cycles and doses administered;
- Number and percentage of subjects with at least 1 dose reduction, number of dose reductions, and reasons (Per Protocol, adverse event, or other) for reduction, by cycle and overall;
- Number and percentage of subjects with at least one dose delay, number of dose delays, by cycle and overall.
- Number and percentage of subjects with at least 1 dose not administered, number of doses not administered, and reasons (adverse event or other) for dose not administered, by cycle and overall. If the reason for dose not administered is not given, it will be categorized as "Data Not Available" in summary tables.

7.3.11. Efficacy Analysis

All efficacy evaluations will be conducted using the ITT population. Supportive analysis of the primary and key secondary efficacy endpoints using the PP and RE population will be conducted for the final analysis. Statistical comparisons will be made between BSC and *nab*-paclitaxel+BSC arms for the Maintenance Part only. Efficacy results that will be considered statistically significant after consideration of the strategy for controlling the Type 1 error rate are described in Section 7.3.11.1. All statistical tests will be two-sided at the significance level of $\alpha = 0.05$, and the corresponding p-values and two-sided confidence intervals (CIs) for intended point estimates will be reported.

Listings will be provided for all endpoints. To support review of tumor response data, the following information will be presented in listings:

- Per subject visit:
  - Date of assessment
  - Number of target lesions
  - Total length of target lesions
  - Percent change from nadir of sum of total length (derived as in Section 7.3.11.2 to review assessment of progressive disease)
  - Percent change from baseline of sum of total length to review assessment of response
  - Assessment of target lesions
  - Assessment of non-target lesions
  - Presence of symptomatic deterioration

- Overall tumor response assessment
- Best overall response

- Per lesion at each subject visit:
  - Location
  - Method of assessment
  - Tumor length (target lesions only)
- Overall per subject
  - The best response over the entire study

7.3.11.1. Multiplicity

A step down procedure will be used to control the family-wise Type 1 error rate. Overall Survival will only be tested if PFS demonstrates superiority for *nab*-paclitaxel+BSC vs. BSC.

Strategies to control alpha due to repeated testing are described in Section 8.2 .

7.3.11.2. Analysis of Primary Efficacy Endpoint

The primary efficacy endpoint of PFS will be analyzed using the p-value from the stratified log-rank test of PFS based on the ITT population and will be considered the primary analysis, with the un-stratified analysis considered a supporting analysis. The method for determining the strata is detailed below.

Progressive disease is measured using the percent change from nadir over all tumor assessments in sum of total length. This is calculated as follows:

Percent change from nadir at a given visit= (total length – nadir in total length) / (nadir in total length) x 100%, where nadir is calculated using all measurements before the visit in question, including baseline, Induction, and Maintenance scans.

Progression-free survival will be based on investigator assessment according to RECIST 1.1 criteria. PFS is defined as the time in months from the date of randomization to the date of disease progression (documented by CT-scan result, not including symptomatic deterioration) or death (any cause) on or prior to the clinical cutoff date, whichever occurs earlier. RECIST 1.1 criteria defines disease progression as the occurrence of any of the following circumstances:

- Increase of 20% in the sum of diameters of the target tumors based on the smallest sum (nadir) over the study (for this study it is possible that the nadir occurs at baseline or during the Induction Part of the study even though those time points are before randomization), with a 5 mm minimum increase
- Appearance of new lesions
- Unequivocal progression of non-target tumors

Therefore, the calculation of PFS is a two-step process:

- Determination of date of disease progression, which is the date of assessment corresponding to the first time that overall tumor response is “Progressive Disease” on the Overall Response eCRF

- Calculation of PFS as (date of disease progression, death, or censoring – date of randomization + 1)/30.4375 as described below

Subjects who do not have disease progression and have not died, regardless of whether they were discontinued from treatment or not, will be censored at the date of last tumor assessment, on or prior to the clinical cutoff date that the subject was progression-free. If a subject begins a subsequent anti-cancer chemotherapy prior to documented disease progression (or death), the subject will be censored at the date of last assessment when the subject was documented as progression-free prior to the intervention. In the event that curative radiotherapy or surgery at lesion sites occurs, the subject will be censored at the date of last assessment when the subject was documented as progression-free prior to the intervention. Subjects with a single missing radiologic assessment prior to a visit with documented disease progression (or death) will be analyzed as a PFS event at the time of the radiologic assessment that shows progression or death (whichever is earlier). Subjects with two or more consecutive missing radiologic assessments prior to an assessment with documented progression (or death) will be censored at the date of last assessment when the subject was documented as progression-free prior to the first of the two missing assessments. This is further illustrated in . below.

Table 5: Censoring Rules for PFS

Situation	Analysis Date	Censored or Event
Disease progression, and time interval between progression date and previous tumor assessment date with progression-free response is less than or equal to **91** days	Earliest of: • Date of first tumor assessment showing new lesion • Date of first tumor assessment with overall response of PD	Event
Disease progression, and time interval between the progression date and the previous tumor assessment date with progression-free response is greater than **91** days, or time interval between the progression date and the randomization date is greater than 91 days if the PD occurred at the first post baseline tumor assessment	Latest of: • the last progression-free assessment date • randomization date	Censored
Death without any post-baseline radiological assessment, and time interval between death date and randomization date is less than or equal to **91** days	Death date	Event

Death before documented progression, and time interval between death date and previous tumor assessment date with progression-free response is less than or equal to **91** days	Death date	Event
Death without any post-baseline radiological assessment, and time interval between death date and randomization date is greater than **91** days	Randomization date	Censored
Death before documented progression, and time interval between death date and previous tumor assessment date with progression-free response is greater than **91** days	The last progression-free assessment date	Censored
No death or disease progression and no subsequent anti-cancer therapy (i.e., systemic anti-cancer therapy, anti-cancer surgery, curative radiotherapy).	Latest of • the last progression-free assessment date • randomization date	Censored
Subsequent anti-cancer treatment (i.e., systemic anti-cancer therapy, anti-cancer surgery, curative radiotherapy) started prior to progression or death	Latest of • the last progression-free assessment date prior to start of anti-cancer treatment • randomization date	Censored
Disease progression and the progression date is before the randomization date	Randomization date	Censored
Note: Progression-free response refers to a response that was neither progressive disease (PD) nor un-evaluable (UE).		

A superiority test will be performed for the primary efficacy endpoint of PFS. The corresponding two-sided hypotheses are:

H_0: HR(*nab*-paclitaxel+BSC vs. BSC only) = 1

versus

H_1: HR(*nab*-paclitaxel+BSC vs. BSC only) $\neq$ 1,

where HR(*nab*-paclitaxel+BSC vs. BSC only) is the hazard ratio between two treatment arms in the Maintenance Part of the study.

The final analysis for PFS will be conducted when a total of approximately 136 PFS events occur. All disease progression and death events that have occurred on or prior to the clinical cutoff date will be included. The survival distribution of PFS will be estimated using the Kaplan-Meier method. The median PFS, including the two-sided 95% confidence interval (CI) for each treatment arm, will be provided. The survival distributions for the two treatment arms will be compared using the stratified log-rank test, and the p-value will be provided (Klein, 2003). The stratification factors are disease stage at diagnosis (IIIB vs. IV), response in Induction (CR/PR vs. SD), and ECOG performance status (0 vs. 1) after Induction. Refer to the pooling of sparse stratums below for situations in which the full stratified analysis will be replaced by a pooled stratified analysis due to a lack of information which could adversely affect power and modeling estimation.

The PFS rates, including two-sided 95% confidence intervals, will be provided for every 2 months relative to randomization (Greenwood's formula, Klein, 2003). Kaplan-Meier curves will be provided by treatment arm. The associated HR and two-sided 95% confidence interval will be estimated by using a stratified Cox proportional hazard model, with the same stratification factors (Cox, 1972). Refer to the pooling of sparse stratums below for situations in which the full stratified analysis will be replaced by a pooled stratified analysis due to a lack of information which could adversely affect power and modeling estimation.

Pooling Algorithm for Eliminating Sparse Stratums

If there exists unacceptable sparseness in some stratum cells for the smallest treatment arm due to the 2:1 randomization ratio (i.e., a stratum cell that has fewer than 5 PFS events for the BSC treatment arms in the ITT population of the Maintenance part), then an algorithm for eliminating some full strata will be employed until there are at least 5 events in any given revised stratum cell. The initial stratum cells for the combined treatment arms will start out as 8 cells that represents all possible permutations of the 3 strata (baseline Eastern Cooperative Oncology Group [ECOG] performance status at the end of the Induction part [0 vs. 1]; tumor response to induction chemotherapy [CR/PR vs. SD]; and disease stage before administration of induction therapy [IIIB vs. IV]). Hence, the 8 cells representing all possible permutations among the strata may be reduced down to either 4 cells (all possible permutations of just 2 relatively large strata), 2 cells (just 1 large stratum); or no strata at all when implementing the algorithm. However, it is extremely unlikely that the algorithm will result in no stratum. Once the revised strata have been determined through this algorithm, then it will be used consistently for all endpoints and populations involving a pre-specified stratified analysis.

If a sparse strata cell exists, then first consider all 3 pooling possibilities to start off with and then select the pooled stratum that provides the largest number of events in the smallest stratum cell. For example, consider three candidate pooling sets (a, b), (a, c), and (b, c), where set (a, b) has event cell counts {(a1, b1), (a1, b2), (a2, b1), (a2, b2)}; set (a, c) has event counts {(a1, c1), (a1, c2), (a2, c1), (a2, c2)}; and set (b, c) has event counts {(b1, c1), (b1, c2), (b2, c1), (b2, c2)}. Now let (a, b)* = min { (ai, bj) } for i=1, 2 and j=1, 2; and (a, c)* = min { (ai, cj) } for i=1, 2 and

j=1, 2; and (b, c)* = min { (bi, cj) } for i=1, 2 and j=1, 2. The newly pooled stratum would be a two strata combination of a, b, or c that is the maximum of set { (a, b)*, (a, c)*, (b, c)* } for i=1, 2 and j=1, 2. In the case of tied cases (e.g., (a, c) and (b, c) are both suggested as the two strata to be used in the analysis), then priority will be based on what is clinically more relevant to keep.

If a sparse strata cell still exists in the first round of pooling, then the above pooling process would start over again to form a single stratum. The pooled candidate stratums are now a, b, and c with the final pooled strata being the max { a*, b*, c* } in which no sparse strata cell exists. In selecting the final pooled strata; a* = min{ ai } for i=1, 2; b* = min{ bi } for i=1,2; and c* = min {ci } for i=1, 2. Again, in the case of tied cases, (e.g., b and c are both suggested as the strata to be used in the analysis), then then priority will be based on what is clinically more relevant to keep.

If there still exists a sparse cell, then the primary analysis will default to an unstratified one.

For this particular application,

a = baseline Eastern Cooperative Oncology Group [ECOG] performance status at the end of the Induction part [a1=0 vs. a2=1] strata;
b = tumor response to induction chemotherapy [b1=CR/PR vs. b2=SD] strata;
c = disease stage before administration of induction therapy [c1=IIIB vs. c2=IV]) strata.

In case of ties, clinical priority will be given to stratums a) baseline ECOG at end of Induction part; b) tumor response to induction chemotherapy; and c) disease stage before administration of induction therapy in this order (i.e., strata a: baseline ECOG has the highest priority and strata c: disease stage before administration of induction therapy has the lowest priority).

Supportive and Sensitivity Analyses

The following analyses will be conducted as supporting or sensitivity analyses using the same methods (i.e. p-value from stratified log-rank test, HR and two-sided 95% confidence interval from the stratified Cox model):

- PFS as calculated from Day 1 of treatment for the Induction Part. For this analysis, the stratification factors will be the same as those used in the primary efficacy analysis.
- PFS where subsequent anti-cancer therapy is considered a PFSevent
- To assess the impact on PFS of response assessments not occurring at the regularly scheduled assessment times, the frequency of these unscheduled/off-scheduled visits will be presented for each treatment arm. In addition, sensitivity analyses will be performed where subjects with events and censorings that occur at a time other than the regularly scheduled visit assessment, will have PFS time based on the date of the next regularly scheduled assessment rather than the actual off-schedule date. The methods similar to the primary analysis will be used to estimate the median PFS along with HRs.

- An additional analysis of PFS will be performed using the European Medicines Agency methodology for analysis of a PFS endpoint. Similar to the approach described above, PFS will be defined as the time from the randomization date to the start of disease progression or subject death (any cause), whichever occurs first. Subjects who do not have disease progression or have not died will be censored at the last known time that the subject is progression free (i.e. the last tumor assessment). However, occasional missing observations or initiation of subsequent anticancer therapy will not result in censoring for this analysis.
- Sensitivity analyses to address the impact of measurability of endpoints and adherence to protocol will be done by performing the PFS analyses mentioned above on the response evaluable and per protocol populations.

7.3.11.3. Analyses of Secondary Efficacy Endpoints

7.3.11.3.1. Overall Survival

Overall survival is defined as the duration in months between randomization and death from any cause:

Time to OS = (Day of death or censoring– Day of randomization + 1)/30.4375

Subjects who are still alive at the end of the study will be followed up for overall survival for 1.5 years after final analysis of approximately 136 PFS events or approximately 147 deaths have occurred, whichever is earlier. Subjects who are still alive after this time will have their overall survival censored at the date of last contact or clinical cutoff, whichever is earlier. The last contact date is the last non-imputed date of any record in the database for that subject, or if the subject is dead, the last known date the subject was alive.

However in the interim analysis of PFS a different rule will be used for simplicity. All subjects who do not have a death record prior to or on the cutoff date will be censored at the 'last date known alive'. For subjects who have withdrawn consent during the study, the last date known alive will be the date of consent withdrawal from the study. For all other subjects, the last date known alive will be the data cutoff date.

Analysis of OS will parallel the analysis of PFS described in Section 7.3.11.2. P-values generated from these analyses will only be interpreted if the primary analysis is significant. The p-value for the stratified log-rank test using the ITT population will be considered the main analysis for the secondary endpoint, with unstratified analyses and analyses based on per protocol and response evaluable populations considered supporting analyses.

Additional supporting and sensitivity analyses will be performed as follows:

- Overall Survival calculated from Day 1 of treatment for the Induction Part (Time to OS = Day of death – Day 1 of treatment for the Induction Part + 1),
- Overall survival using the additional censoring rule of Day 1 of treatment for the cross-over part for subjects in BSC arm who crossed over to *nab*-paclitaxel before protocol amendment 2 was implemented.

To assess the impact of starting subsequent anticancer therapy on subject survival, a sensitivity analysis will be conducted. Subjects who start a subsequent anti-cancer therapy will be censored

at the initiation date of the subsequent chemotherapy; a method similar to the primary analysis will be applied to estimate the medians of OS and compare the survival distributions between two treatment arms.

7.3.11.3.2. Tumor Response and Overall Response Rate

Overall response over the entire study is defined as a complete or partial response (CR/PR) to BSC or *nab*-paclitaxel+BSC based on the baseline established within 28 days on or before the start of Induction therapy according to RECIST 1.1 criteria and confirmed in no less than 28 days. Thus, a subject will be considered to have an overall response if there is any tumor assessment where the confirmed overall tumor response is "Complete Response" or "Partial Response." The best response among any tumor assessments after Day 1 of treatment for the Induction treatment up until the subject receives a subsequent anti-cancer therapy will be used for this analysis. This will be calculated for randomized subjects only.

The ORR will be summarized by number and percent of randomized subjects with CR/PR and presented with 95% confidence intervals for the percent. The treatment effect size will be estimated by the response rate ratio ($P_{nab\text{-}P+BSC}/P_{BSC}$). The ORR will be analyzed using the Cochran-Mantel-Haenszel (CMH) test with stratification factors:

- Disease stage at diagnosis (IIIB vs. IV);
- Response in Induction (CR/PR vs. SD);
- ECOG performance status at end of Induction (0 vs. 1).

If strata cell sizes are too small based on Mantel-Fleiss criterion, Pearson chi-square test will be used instead. This analysis based on the ITT population will be used as the main analysis of ORR.

As sensitivity analyses, a CMH (or Pearson chi-square) test with the above strata will be performed on the PP and response evaluable populations.

An additional analysis will be performed on the ITT population of ORR in which subjects with a best overall response of CR or PR with or without confirmation of response are considered responders.

The number and percentage of those with a best overall tumor response over the entire study of PD, SD, PR, CR, and unevaluable (UE) will be presented. The denominator for the percentage will include all ITT subjects (subjects who were randomized in the Maintenance Part), regardless of whether they had a tumor assessment. Subjects with no tumor assessments will be considered UE. This analysis will be presented for each protocol-specified CT scan. If there is more than one CT scan within a relative day range, the best response will be counted. Relative study day windows are presented in Section 11.4.

The best (i.e., smallest) percent change from baseline will be calculated for all subjects and will be presented by treatment arm in a waterfall plot.

7.3.11.4. Analyses of Other Secondary Endpoints

7.3.11.4.1. Disease Control Rate

Disease Control rate is defined as the percent of subjects who have a radiologic complete response, partial response, or stable disease according to RECIST 1.1 guidelines, as determined by the investigator. Disease Control rate over the entire study along with associated Clopper-Pearson 95% CI will be presented for the Maintenance ITT population. The relative treatment effect will be summarized by the ratio of the disease control rate and the associated two-sided 95% CI.

7.3.11.4.2. Time to Response

Time to confirmed response is defined as the time from the first day of Induction study drug to the first occurrence of confirmed response (CR or PR) over the entire study. The time to confirmed response will be presented by treatment arm using the summary statistics. Only subjects with a confirmed CR or PR as a best overall response will be included in this analysis.

The same analysis will be performed for time to response with or without confirmation of response.

7.3.11.4.3. Duration of Response

For subjects who had a confirmed CR or PR, the duration of overall response is measured from the time measurement criteria are first met for CR/PR (whichever is first recorded) until the first date that recurrent or progressive disease is radiologically documented (taking as reference for progressive disease the smallest measurements recorded on study).

Subjects who are non-responders (i.e., do not achieve at least a PR) will be excluded from this analysis. Subjects who do not have PD after the response will be censored on the date of last tumor assessment. If a subject died before PD, then the subject will be censored on the date of death.

The duration of response will be analyzed using the Kaplan-Meier method. The median time (including 2-sided 95% CI) will be summarized for each treatment arm; the associated hazard ratio with two-sided 95% confidence interval will be estimated using a stratified Cox proportional hazard model (Cox, 1972). For this analysis, the stratification factors will be the same as those used in the primary efficacy analysis, if stratum sizes are sufficient.

The same analysis will be performed for duration of response with or without confirmation of response.

7.3.12. Subgroup Analyses

PFS, OS, and ORR will all be analyzed within the following subgroups (with subgroup data based on the clinical database rather than IVRS). Subgroup factors will not be used as regression adjustment covariates for relevant analyses, but otherwise these analyses will parallel the analyses for the endpoint performed on the whole population:

1. Age (< 65 years and ≥ 65 years, < 70 years and ≥ 70 years, < 75 years and ≥ 75 years);
2. Race (white versus non-white)

3. ECOG status (0 versus 1) at end of Induction;
4. Sex (male versus female);
5. Region (North America versus Rest of the World)
6. Stage at diagnosis (IIIB versus IV);
7. Response at the end of Induction (PR/CR versus SD);
8. Subjects with medical history that includes diabetes

If there are less than 5 PFS events per treatment per subgroup for the smaller treatment subgroup, we will drop the subgroup analysis for PFS, OS and ORR.

Forest plots will be provided displaying the HRs and corresponding 95% CIs for each subgroup for PFS and OS. The overall HRs with 95% CIs will also be provided in forest plots.

7.3.13. Analyses of Exploratory Efficacy Endpoints

7.3.13.1. Overall Response Rate in Maintenance Part beyond Response in Induction

Overall response during the Maintenance Part beyond the response to Induction therapy is defined as a complete or partial response (CR/PR) based on RECIST 1.1, taking as reference the sum of the longest diameters of the target lesions measured at the last tumor assessment before randomization. For example, if the sum of the longest diameters of the target lesions at the last CT-scan prior to randomization is 10 mm, then the subject will have to have a ≥3 mm reduction from the 10 mm during the Maintenance Part in order for the subject to be deemed having partial response, assuming the criteria for the non-target lesions and new lesions are satisfied also. Naturally, only subjects who have partial response or stable disease at the end of the Induction Part will be able to have further improvement during the Maintenance Part according to this definition.

The target lesions identified prior to the start of Induction therapy will be retained as the target lesions for this exploratory analysis. All subjects who enter the Maintenance Part in each treatment arm will be counted in the denominator of this analysis.

The percentage of subjects achieving partial response, complete response, and either partial or complete response (ORR) as defined above will each be presented with the associated 95% confidence interval.

7.3.13.2. Tumor Characteristics

The following analyses will be covered under a separate SAP:

- The correlation between pretreatment tumor characteristics and response and resistance to the study treatment determined using next-generation sequencing methods, immunohistochemistry, or other analysis methods.
- The association between the changes in tumor characteristics and the acquisition of resistance to therapy from plasma samples taken at treatment failure during Maintenance.

- The relationships between variants in subject pharmacogenomics, such as polymorphisms in enzymes involved in drug metabolism, and treatment efficacy and toxicity.

7.3.13.3. Healthcare Resource Utilization

Summary statistics for location of care (number of office visits, hospital outpatient visits, hospital inpatient visits, emergency room visits, home healthcare visits, hospice, unknown, and other utilizations) and provider (general physician, specialist, nurse practitioner/physician assistant, paramedical care, unknown, and other) will be presented by treatment arm at each time point the utilization questionnaire is given (i.e. Day 1 of every cycle and Early Treatment Discontinuation/End of Cycle 4 during Induction, Day 1 of every cycle and disease progression during Maintenance Part of study).

7.3.14. Safety Analysis

The purpose of this section is to define the safety parameters for the Maintenance Part. All summaries of safety data will be conducted using the safety population.

7.3.14.1. Adverse Events

Adverse events will be analyzed in terms of treatment-emergent adverse events (TEAEs), which are defined as any AEs that begin or worsen after the randomization date and through 28 days after the treatment discontinuation date. In addition, any SAE with an onset date more than 28 days after treatment discontinuation date that is assessed by the investigator as related to study drug will be considered a TEAE. All AEs will be coded using the Medical Dictionary for Regulatory Affairs® (MedDRA) dictionary Version 17.0 or higher. The severity will be graded based on NCI CTCAE Version 4.0.

The following rules will be implemented for cycle calculations:

- TEAEs which start after randomization date and on or before Cycle 6 Day 1 belong to only Cycle 5.
- TEAEs will be categorized by the "throw-back rule", that is, AEs that start on Day 1 of a cycle will be allocated to only the previous cycle.
- All TEAEs which start after Day 1 of the last cycle will be included only in the last cycle.

For the summary of treatment related AEs, a treatment-related TEAE is defined as an adverse event which was considered to be related, reported as "suspected" in eCRF, to *nab*-paclitaxel. If a subject experiences multiple occurrences of the same AE with different relationship to study medication categories, the subject will be counted once, as a relationship category of treatment related. AEs with a missing relationship will be presented in the summary table as a relationship category of "treatment-related".

The summaries below will be presented for the Maintenance Part of the study and based on the safety population.

If a subject experiences the same AE more than once with different toxicity grade, then the event with the highest grade will be tabulated in "by grade" tables. If a subject experiences multiple

AEs under the same preferred term (system organ class), then the subject will be counted only once for that preferred term (system organ class). In addition, AEs with a missing intensity will be presented in the summary table as an intensity category of "Missing."

The incidence of TEAEs will be summarized by MedDRA system organ class (SOC) and preferred term (PT). The intensity of AEs will be graded 1 to 5 according to the National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE) Version 4.0. For all other AEs not described in the CTCAE criteria, the intensity will be assessed by the investigator as mild (grade 1), moderate (grade 2), severe (grade 3), life-threatening (grade 4) or death (grade 5). Tables summarizing the incidence of TEAEs by treatment arm will be generated for each of the following:

- Overall summary of TEAEs;
- TEAEs presented by MedDRA system organ class, preferred term and worst CTCAE grade;
- Treatment-related TEAEs by grade category (grade 1-2 vs. grade 3-4 vs. grade 5; overall and by cycle);
- Serious TEAE by worst CTCAE grade;
- Treatment-related serious TEAE by worst CTCAE grade;
- TEAEs with action of study drug withdrawn;
- Treatment-related TEAEs with action of study drug withdrawn;
- TEAEs with action of study drug dose reduced or interrupted;
- Treatment-related TEAEs with action of study drug dose reduced or interrupted;
- TEAEs that lead to death;
- Treatment-related TEAEs with fatal outcome;
- All death within 28 days of treatment discontinuation date with cause of death;
- Most frequent TEAEs (≥ 5% in either arm; sorted by preferred term, without SOC)
- TEAEs for the following baseline subgroups (provided the number of subjects are sufficient):
  - Age (<65 versus 65 to ≤ 74 years versus ≥ 75 years)
  - Sex (Male versus Female)

All TEAEs will be listed. A listing for non-treatment-emergent AEs will also be provided.

7.3.14.2. Peripheral Neuropathy

Peripheral neuropathy events will be collected in the Physician Assessment of Peripheral Neuropathy eCRF and reported as AEs. PN AEs will be included in analyses described in Section 7.3.14.1. A summary of PN grade by treatment arm and cycle will be provided based on the eCRF data. In addition, time from first dose in Induction to the first occurrence or worsening

of peripheral neuropathy and time from occurrence or worsening to improvement in peripheral neuropathy will be presented by treatment arm using both the AE data and the PN eCRF data as follows:

- Time to first onset of grade 3 or higher peripheral neuropathy;
- Time to improvement of grade 3 or higher peripheral neuropathy by at least 1 grade; and
- Time to improvement of grade 3 or higher peripheral neuropathy to grade 0 or grade 1.

Due to the small number of events expected, time to first onset will be summarized with descriptive statistics for only those subjects who experienced peripheral neuropathy of the relevant grade. There will be no censoring.

Time to improvement will include only subjects who experienced peripheral neuropathy of the relevant grades. Subjects who do not experience improvement will be censored at the last time the subject is evaluated for adverse events. Kaplan Meier methods will be used to attain the median and 95% confidence interval.

7.3.14.3. Adverse Events of Special Interest

The following summaries will be provided for TEAEs included in the above-mentioned AEs of interest:

1. TEAEs by maximum CTCAE grade;
2. Serious TEAEs;
3. TEAEs with action of treatment discontinuation;
4. TEAEs with action of dose reduction or interruption;
5. TEAEs with fatal outcome;

In addition, time from first dose in Induction to the first occurrence or worsening of Myelosuppression and Peripheral Neuropathy and time from occurrence or worsening to improvement in Myelosuppression and Peripheral Neuropathy will be presented by treatment arm as follows:

- Time to first onset of grade 3 or AESI;
- Time to improvement of grade 3 or AESI by at least 1 grade; and
- Time to improvement of grade 3 or higher AESI to grade 0 or grade 1.

Due to the small number of events expected, time to first onset will be summarized with descriptive statistics for only those subjects who experienced an AESI of the relevant grade. There will be no censoring.

Time to improvement will include only subjects who experienced an AESI of the relevant grades. Subjects who do not experience improvement will be censored at the last time the subject

is evaluated for adverse events. Kaplan Meier methods will be used to attain the median and 95% confidence interval.

AESIs will be listed.

7.3.14.4. Clinical Laboratory Evaluations

7.3.14.4.1. Hematology

All hematology data will be listed.

7.3.14.4.2. Clinical Chemistry

All clinical chemistry data will be listed.

7.3.15. Quality of Life Analysis

The LCSS and EQ-5D-5L, as described in Section 7.2.12 , will be administered in the Maintenance Part of the study as well. Baseline scores are defined as scores captured on Day 1 of Induction.

7.3.15.1. LCSS

LCSS scores (overall constitutional score, symptom burden index, respiratory symptom scale, 3-item scale) will be calculated in the same way as described in Section 7.2.12.1.

Summary statistics of the individual LCSS items and summary scale scores, the change from baseline, and percent change from baseline for these scores will be summarized by treatment arm for the ITT population at every Maintenance Part time point at which the instrument is given, i.e. Day 1 of every cycle. In addition, the change from baseline to the LCSS score during the Maintenance Part, defined as the difference between the last reported LCSS score during the Maintenance Part, last visit or early discontinuation visit or the last reported LCSS score before clinical cutoff for subjects who have not reached that point at an interim analysis, will be presented.

Statistical comparisons will be made between the BSC and *nab*-paclitaxel+BSC groups in the Maintenance Part for change from baseline to last administration in Maintenance using analysis of covariance. Disease stage, ECOG status, and response during Induction will be the covariates.

7.3.15.2. EQ-5D-5L

EQ-5D scores will be calculated in the same way as described for Baseline scores and are defined as the score captured on Day 1 of Induction, and missing baseline scores will not be imputed.

The utility score as well as the change from baseline and percent change from baseline in the utility score will be summarized for every time point at which the instrument is given for the ITT population. In addition, the change from baseline to the utility score during the Induction Part, defined as the difference between the last reported overall score during the Induction Part, early discontinuation or disease progression visit or the last reported overall score before clinical cutoff for subjects who have not reached that point at an interim analysis will be presented.

Summaries of the VAS (with the exception of the missing values) results, the change from baseline, and percent change from baseline will be presented for every time point at which the instrument is given. In addition, the change from baseline to the last reported overall score during the Maintenance Part, either the disease progression or early discontinuation visit or the last reported overall score before clinical cutoff for subjects who have not reached that point at an interim analysis will be presented.

Shifts from baseline to each time point at which the instrument is given will be presented for each dimension of the EQ-5D-5L for the ITT population.

For the utility score and the VAS, statistical comparisons will be made between the BSC and *nab*-paclitaxel+BSC groups in the Maintenance Part for change from baseline to last administration in Maintenance using analysis of covariance. Disease stage, ECOG status, and response during Induction will be the covariates.

7.4. Follow-up

First subsequent systemic anti-cancer regimens will be summarized by treatment arm.

To assess how quickly the second line therapy will be initiated for the subjects who discontinued the study treatment, the time to first subsequent therapy, defined as the time interval from the date of treatment discontinuation to the start date of the first subsequent anti-cancer regimen. The start date of the first subsequent anti-cancer regimen is the earliest start date of any anti-cancer therapy taken in the follow-up/survival period. Subjects who don't have a subsequent therapy and have not died as of the cutoff date for the statistical analysis will be censored at the last follow up date. Subjects who die before initiating a subsequent therapy will be censored at the date of death.

The time to second line therapy will be analyzed using the Kaplan-Meier method. The median time (including 2-sided 95% CI) will be summarized for each treatment arm; the associated hazard ratio with two-sided 95% confidence interval will be estimated using the Cox proportional hazard model (Cox, 1972).

In addition, first subsequent anti-cancer treatment regimens will be summarized by the number and percentage of subjects receiving each regimen category and regimen; the number of cycles overall, and the number of cycles for each regimen category and regimen administered. The best response overall and for each regimen category and regimen administered will be summarized. Treatment duration, defined as (date of last dose of any treatment in a regimen - date of first dose of any treatment in the same regimen) + 1 will be summarized as well. Additional subsequent anti-cancer treatment regimens will be listed.

Anti-cancer surgeries during follow-up will be summarized by treatment group and by system organ class and preferred term Follow-up surgeries will be listed.

Radiation therapy during follow-up will be summarized by treatment arm and by category (prior, concomitant, follow up); type (External beam, radio-immuno therapy, brachytherapy, other); location (if external beam); dose; fraction; intent (adjuvant, curative, palliative, unknown); and setting (stand alone radiation therapy, concurrent with other anti-cancer therapy, sequential to other anti-cancer therapy). Follow-up radiation therapy will be listed.

8. INTERIM ANALYSIS

8.1. General Information

An independent DMC will be convened with experts not otherwise involved in the study as investigators. During the course of the study, the DMC will review the progression-free survival efficacy data once and overall survival data separately in accordance with the guidelines for the pre-planned interim analysis. The committee will also review safety data periodically. Operational details for the DMC and the algorithm and its validation by an expert panel will be detailed in the DMC charter.

The DMC is advisory in nature, and the decision to implement DMC recommendations will be made by the sponsor. Specifically, the futility analyses for PFS and OS are nonbinding.

For data cleaning purposes, a cut-off date of December 12, 2016 in conjunction with the cut-off date algorithm will be used to create the necessary analysis populations for the interim analysis. The details of data cut-off algorithm for the interim analysis will be provided in a separate file. In order to effectively deal with a substantial number of over-runs expected to occur after this December 12th cut-off date, the most current PFS data will be analyzed again using the Cox regression model and a plot of Kaplan-Meier product limit estimates. All data for patients not experiencing a PFS event may not be totally clean as of the December 12th cut-off date.

8.2. Statistical Approaches for Control of Alpha

There will be one non-binding interim analysis for futility for the primary endpoint, PFS, when approximately 91 events (67% information) have been observed. Futility may be declared and the study discontinued if the conditional power at the interim analysis is < 10%. The final analysis of PFS will be performed when approximately 136 PFS events have been observed.

A Step-down procedure from PFS to OS will be used to control the family-wise two-sided Type-I error rate to 5%. That is, OS will be tested only if the *nab*-paclitaxel plus BSC regimen demonstrates superiority over the BSC alone with PFS.

The marginal significance level for OS will be adjusted according to O'Brien-Fleming type of boundary. The nominal two-sided p-values for declaring superiority will be calculated using the statistical software The analysis and boundaries will be adjusted based on the actual number of deaths observed.

There will be two analyses for the secondary OS endpoint. The interim analysis of OS will be conducted based on the number of deaths observed up to the clinical cutoff date. Futility may be declared and the study discontinued if the conditional power at the interim analysis for OS is < 10% or the *nab*-paclitaxel plus BSC regimen does not demonstrate superiority over the BSC alone with PFS. For efficacy, the two-sided nominal value p-value stopping boundary for OS interim analysis is 0.009. Efficacy will be declared and study discontinued if p-value crosses stopping boundary and the *nab*-paclitaxel plus BSC regimen demonstrates superiority over the BSC alone with PFS. If neither futility nor efficacy is declared at the interim analysis of OS, the

final OS analysis will be performed after approximately 147 deaths have been observed, if applicable. The final two-sided nominal p-value efficacy stopping boundary is 0.047.

8.3. Probability of Clinically Meaningful PFS Effects Given the Interim Data

At the interim analysis, the probability of clinically meaning PFS effects given the interim data will be provided. Clinically meaningful PFS effects is defined as Hazard Ratio (HR) <0.85.

In the stratified Cox proportional hazard model, denote the Log(Hazard Ratio) parameter as β. The estimated Log(HR) at the interim analysis approximately has a normal distribution $\hat{\beta} \sim N(\beta, \sigma_{\hat{\beta}}^2)$, where $\hat{\beta}$ is the estimator for the hazard ratio (HR) on the natural log scale with mean β and variance $\sigma_{\hat{\beta}}^2$. So the probability of clinically meaningful PFS effects is

$$\Pr\left(\hat{\beta} < Log(0.85)\right) = 1 - \Phi(Z),$$

where $Z = (b - Log(0.85))/SE(b)$ and b is the estimated hazard ratio on the natural log scale with standard error $SE(b)$. b and $SE(b)$ are readily available on SAS Proc PHREG output. The hazard ratio of 0.85 is deemed to be the minimal clinically significant effect. Other probabilities of meaningful HRs that may represent stronger PFS effects will also be provided (e.g., $\Pr(\hat{\beta} < \log(0.80)), \Pr(\hat{\beta} < \log(0.75)), \Pr(\hat{\beta} < \log(0.70)), \Pr(\hat{\beta} < \log(0.60))$). .

8.4. Conditional Power at the Interim Analysis for PFS

The two-sided conditional power for PFS at the interim analysis is computed as

$$P_k(\theta) = \Phi\left(\frac{Z_k\sqrt{I_k} - z_{\frac{\alpha}{2}}\sqrt{I_K} + \theta(I_K - I_k)}{\sqrt{I_K - I_k}}\right) + \Phi\left(\frac{-Z_k\sqrt{I_k} - z_{\frac{\alpha}{2}}\sqrt{I_K} - \theta(I_K - I_k)}{\sqrt{I_K - I_k}}\right)$$

(Jennison and Turnbull (2000) Page 205-208)

where

$\theta = \log(HR) = \log\left(\frac{\lambda_{nab+BSC}}{\lambda_{BSC}}\right) = \log(0.6)$ (the log hazard ratio under the alternative hypothesis);

Z_k: the unstratified log-rank test statistic computed from the observed data at the interim;

I_k: the information level at the interim stage, $I_k = E_k P_1(1 - P_1)$;

E_k: the number of PFS events at the interim stage;

P_1: the proportion of the subjects randomized to nab+BSC arm;

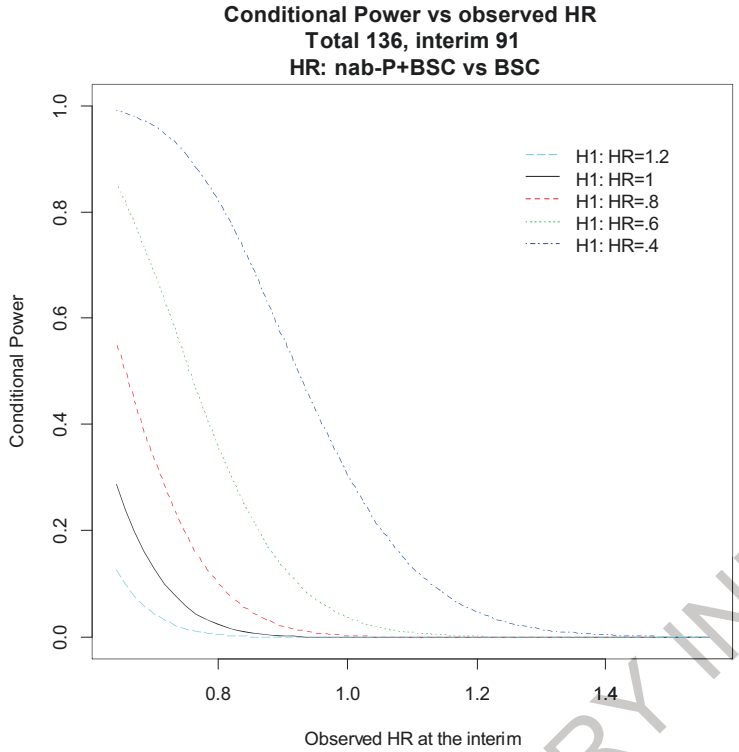
I_K: the information level at the end of the study, $I_K = EP_1(1 - P_1)$;

E: the total number of events at the end of study (136);

$z_{\frac{\alpha}{2}}$: 1.96.

More details about sequential trial design can be found in Lan and DeMets 1983; Lan and Zucker 1993; Jennison and Turnbull 2000.

Conditional Power depends on the observed test statistic Z_k and the alternative hypothesis θ. The following graph shows the relationship between the observed hazard ratio and the conditional power at the interim analysis under different assumptions of alternative hypothesis.



The plot above shows the relationship of the conditional power and the observed hazard ratio at the interim based upon the assumed unkown true hazard ratio. The light green hashed line is the assumed true hazard ratio of 0.60 which was used in determining the 136 PFS events required to provide 80% power under a two sided test at the nominal 5% level of significance).

Operating charatceristics for stopping the study at the interim for futility under some various conditional power/ observed hazard ratios based upon 5,000 simulated trials under the protocol assumption of a true hazard ratio of 0.60 are provided in the tables below.

Observed HR at Interim	Conditional Power based on the observed HR and the assumption that the true HR=0.6
1.00	3.7%
0.98	5%
0.94	10%
0.89	15%

Futility Criteria	Percentage of declaring futility at Interim given the true HR=0.6
HR>1.00	1.0%
HR>0.98	1.3%
HR>0.94	2.1%
HR>0.89	3.7%

8.5. Conditional Power at the Interim Analysis for OS

Interim Analysis for OS will be conducted when 136 PFS events are achieved and the *nab*-paclitaxel plus BSC regimen demonstrates superiority over the BSC alone with PFS. Otherwise, OS will not be tested and therefore there will not be an interim analysis for OS.

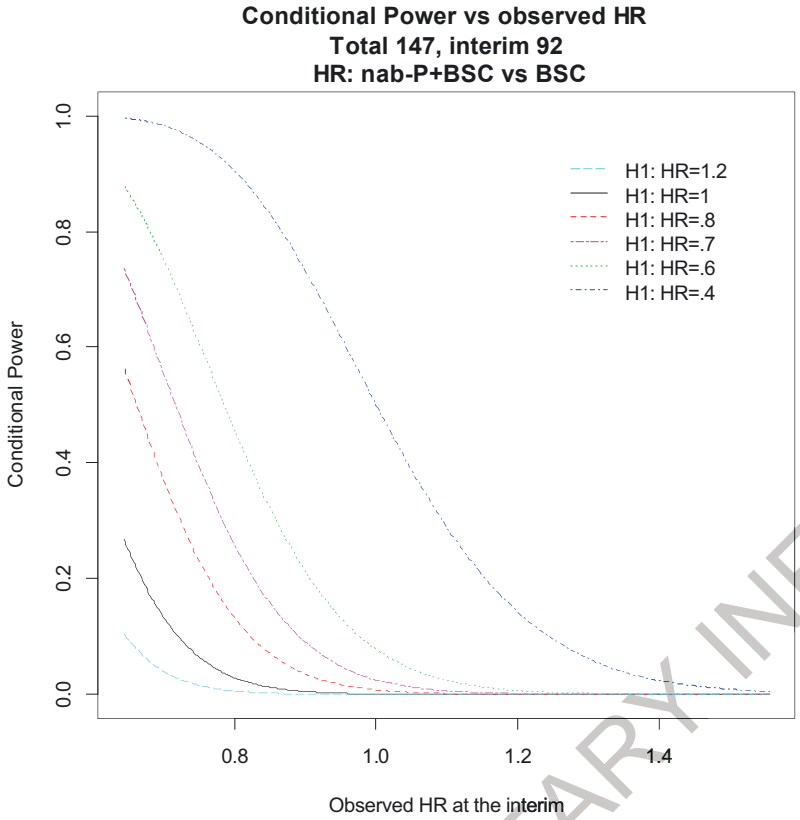
The two-sided conditional power for OS at the interim analysis is computed as

$$P_k(\theta) = \Phi\left(\frac{Z_k\sqrt{I_k} - z_{\frac{\alpha}{2}}\sqrt{I_K} + \theta(I_K - I_k)}{\sqrt{I_K - I_k}}\right) + \Phi\left(\frac{-Z_k\sqrt{I_k} - z_{\frac{\alpha}{2}}\sqrt{I_K} - \theta(I_K - I_k)}{\sqrt{I_K - I_k}}\right)$$

(Jennison and Turnbull (2000) Page 205-208)
where

$\theta = \log(HR) = \log\left(\frac{\lambda_{nab+BSC}}{\lambda_{BSC}}\right) = \log(0.7)$ (the log hazard ratio under the alternative hypothesis for OS);
Z_k: the unstratified log-rank test statistic of OS computed from the observed data at the interim;
I_k: the information level at the interim stage, $I_k = E_k P_1(1-P_1)$;
E_k: the number of OS events at the interim stage;
P_1: the proportion of the subjects randomized to nab+BSC arm;
I_K: the information level at the end of the study, $I_K = EP_1(1-P_1)$;
E: the total number of OS events 1.5 years after final analysis of PFS (147);
$z_{\frac{\alpha}{2}}$: 1.96.

Conditional Power depends on the observed test statistic Z_k and the alternative hypothesis θ. The following graph shows the relationship between the observed hazard ratio and the conditional power at the interim analysis of OS under different assumptions of alternative hypothesis, in which the assumption is that the number of OS events in the interim analysis is 92.



The plot above shows the relationship of the conditional power and the observed hazard ratio at the interim based upon the assumed unkown true hazard ratio. The purple hashed line is the assumed true hazard ratio of 0.70 which was used in determining the 147 OS events required to provide 53% power under a two sided test at the nominal 5% level of significance).

Operating charatceristics for stopping the study at the interim for futility under some various conditional power/ observed hazard ratios based upon 5,000 simulated trials under the protocol assumption of a true hazard ratio of 0.70 are provided in the tables below.

Observed HR at Interim	Conditional Power based on the observed HR and the assumption that the true HR=0.7
1.00	2.5%
0.95	5%
0.89	10%
0.85	15%

Futility Criteria	Percentage of declaring futility at Interim given the true HR=0.7
HR>1.00	5.2%
HR>0.95	8.0%
HR>0.89	13.4%
HR>0.85	18.3%

9. CHANGES TO THE STATISTICAL ANALYSES SECTION OF THE PROTOCOL

For the Induction Part, the following changes were made to the secondary efficacy endpoints:

- Disease control rate: protocol originally stated DCR over the entire study, statistical analysis plan is adding text to clarify DCR will also be analyzed separately during Induction Part

For the Maintenance Part, the following changes were made to the secondary efficacy endpoints:

- PFS from Day 1 Cycle 1, protocol originally stated PFS from randomization into the Maintenance Part of the study, statistical analysis plan is adding additional analyses to further characterize the PFS endpoint;
- OS from Day 1 Cycle 1, protocol originally stated OS from randomization into the Maintenance Part of the study, statistical analysis plan is adding additional analyses to further characterize the PFS endpoint;
- ORR without confirmation required for PR or CR: protocol originally stated ORR during Induction and Maintenance Parts, statistical analysis plan is adding additional analyses to further characterize the ORR endpoint.

10. REFERENCES

Cox, DR. Regression Models and Life-Tables. Journal of the Royal Statistical Society. Series B (Methodological). 1972, 34 (2): 187–220.

Klein JP and Moeschberger ML. Survival Analysis: Techniques for Censored and Truncated Data, 2nd ed. Springer, 2003.

Mantel N and Fleiss JL (1980). Minimum Expected Cell Size Requirements for the Mantel-Haenszel One-Degree-of-Freedom Chi-Square Test and a Related Rapid Procedure. American Journal of Epidemiology, 112: 129-134.

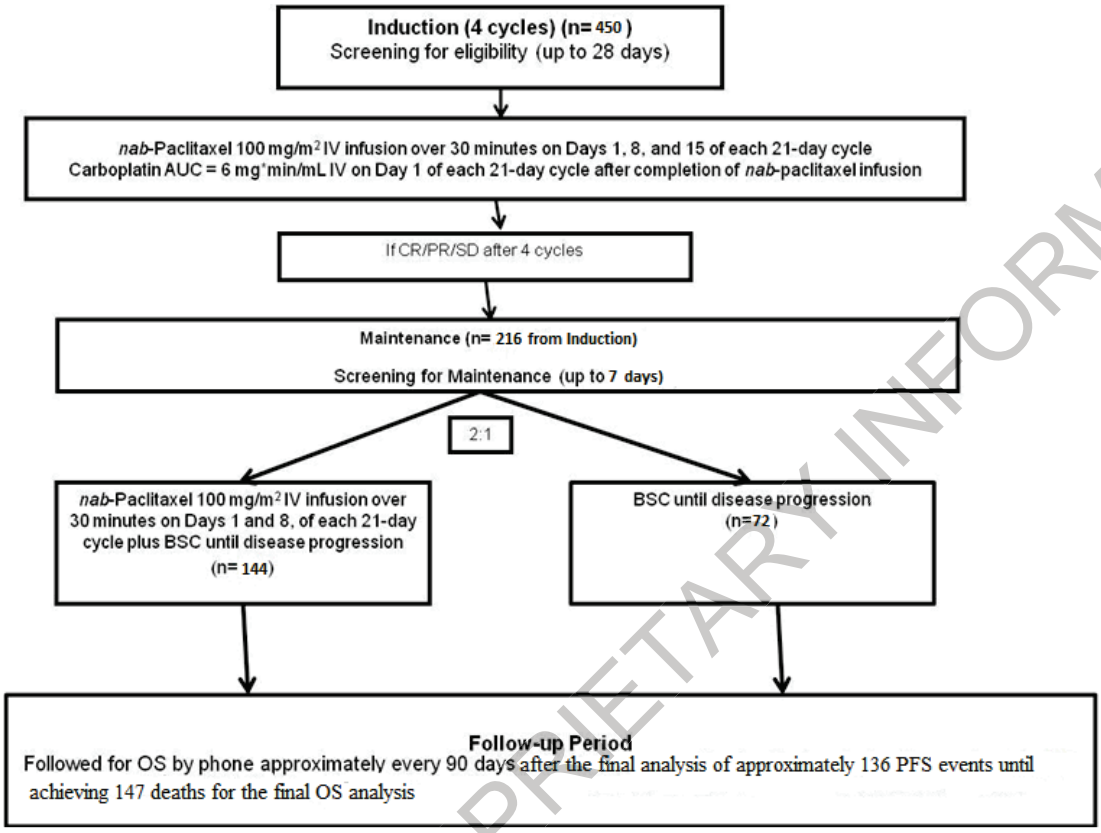
Lan and DeMets (1983). Discrete Sequential Boundaries for Clinical Trials. Biometrika (1983) 70 (3): 659-663.

Lan and Zucker (1993). Sequential Monitoring of Clinical Trials: The Role of Information and Brownian Motion. Statistics in Medicine 12(8):753 - 765

Jennison and Turnbull (2000). Group Sequential Methods with Applications to Clinical Trials. CRC/Chapman & Hall, U.K., 2000.

11. APPENDICES

11.1. Study schematic



11.2. Handling of Dates

Dates will be stored as numeric variables in the SAS analysis files and reported in DDMMMYYYY format (i.e., the Date9. datetime format in SAS). Dates in the clinical database are classified into the categories of procedure dates, log dates, milestone dates, outcome dates, and special dates.

- **Procedure Dates** are the dates on which given protocol-specified procedure are performed. They include the dates of laboratory testing, physical examinations, tumor scans, etc. They should be present whenever data for a protocol-specified procedure are present and should only be missing when a procedure are marked as NOT DONE in the database. Procedure dates will not be imputed.
- **Log Dates** are dates recorded in CRF data logs. Specifically, they are the start and end dates for adverse events and concomitant medications/procedures. They should not be missing unless an event or medication is marked as *ongoing* in the database.

Otherwise, incomplete log dates will be imputed according to the rules in Appendix 16.2 (eg, for duration or cycle assignment, etc). However, in listings, log dates will be shown as recorded without imputation.

- **Milestone Dates** are dates of protocol milestones such as randomization, study drug start date, study drug termination date, study closure date, etc. They should not be missing if the milestone occurs for a subject. They will not be imputed.
- **Outcome Dates** are dates corresponding to study endpoints such as survival, progression, etc. In most cases they are derived either from a milestone (eg, the survival date is derived from the death date), or a procedure date (eg, the progression date is derived from the date of the tumor scan that was used to determine progression). They may be subject to endpoint-specific censoring rules if the outcome did not occur, but are not otherwise subject to imputation.
- **Special Dates** cannot be classified in any of the above categories and they include the date of birth. They may be subject to variable-specific censoring and imputation rules.
- **Last Contact Dates** for the survival analysis are the maximum date collected in the database, if the imputed date used for response date or AE date, the last contact dates should be the latest date of those imputed date and maximum date in the database.

Dates recorded in comment fields will not be imputed or reported in any specific format.

11.2.1. Calculation Using Dates

Calculations using dates (*e.g.*, subject's age or relative day after the first dose of study drug) will adhere to the following conventions:

- Study days after the start day of study drug will be calculated as the difference between the date of interest and the first date of dosing of study drug plus 1 day. The generalized calculation algorithm for relative day is the following:
  - If TARGET DATE >= DSTART then STUDY DAY = (TARGET DATE – DSTART) + 1;
  - Else use STUDY DAY = TARGET DATE – DSTART.

  Note that Study Day 1 is the first day of treatment of study drug. Negative study days are reflective of observations obtained during the baseline/screening period. Note: Partial dates for the first study drug are not imputed in general. All effort should be made to avoid incomplete study drug start dates.
- Age (expressed in days) is calculated: AGE = CONSENT DATE – DATE of BIRTH + 1. In practice, age will be transformed to years by dividing the difference by 365.25 days, then truncating.
  - Preference is for using calculated age from clinical database. When not available, calculated age from CRF or IVRS may be used
  - Partial birth date: impute missing day as 15th of the month; impute missing month as July; set missing age for missing year

- Intervals that are presented in weeks will be transformed from days to weeks by using (without truncation) the following conversion formula:

  WEEKS = DAYS /7

- Intervals that are presented in months will be transformed from days to months by using (without truncation) the following conversion formula:

  MONTHS = DAYS /30.4375

11.3. Date Imputation Guideline

11.3.1. Impute Missing Dates for Adverse Events/ Prior or Concomitant Medications

Incomplete Start Date

- If the stop date is not missing, and the imputed start date is after the stop date, the start date will be imputed by the stop date.

Missing day and month

- If the year is the **same** as the year of the first dosing date, then the day and month of the first dosing date will be assigned to the missing fields.
- If the year is **prior to** the year of first dosing date, then December 31 will be assigned to the missing fields.
- If the year is **after** the year of first dosing, then January 1 will be assigned to the missing fields.

Missing day only

- If the month and year are the **same** as the year and month of first dosing date, then the first doing date will be assigned to the missing day.
- If either the year of the partial date is **before** the year of the first dosing date or the years of the partial date and the first dosing date are the same but the month of partial date is **before** the month of the first dosing date, then the last day of the month will be assigned to the missing day.
- If either the year of the partial date is **after** the year of the first dosing date or the years of the partial date and the first dose date are the same but the month of partial date is **after** the month of the first dosing date, then the first day of the month will be assigned to the missing day.

Missing day, month, and year

- No imputation is needed, the corresponding AE will be included as TEAE if end date of AE is after the first dose date or the end date is also missing.

Incomplete End Date

Missing day and month

- December 31 will be assigned to the missing fields.

Missing day only

- The last day of the month will be assigned to the missing day.

11.3.2. Impute Missing Dates for Medical History

Partially missing medical history start dates will be imputed in the derived dataset for medical history. The 16th of the month will be used to impute a partially missing start date that has only the day missing, and July 1st will be used to impute a partially missing start date that has both the month and day missing.

11.4. Tumor Response Windows

Relative Day Ranges for the CT-Scans

Study Week	Study Part	Relative Weeks in Study Part	Relative Days in Study Part
6	Induction	6	2 - 63
12		12	≥64 and before randomization
18	Maintenance	6	2 - 63
24		12	64 - 105
30		18	106 - 147
36		24	148 - 189
42		30	190 - 231
48		36	232 - 273
54		42	274 - 315
60		48	316 - 357
66		54	358 - 399
72		60	400 - 441
78		66	442 - 483
84		72	484 - 525
90		78	526 - 567
96		84	568 - 609
102		90	610 - 651
108		96	652 - 693
114		102	694 - 735
120		108	736 - 777
126		114	778 - 819
132		120	820 - 861
138		126	862 - 903

11.5. MedDRA Versions and Dates

For reports dated:	MedDRA version
05May2014 – 01Nov2014	17.0
02Nov2014 – 03May2015	17.1
04May2015 – 01Nov2015	18.0
02Nov2015 – present	18.1