1.0 Title Page

Statistical Analysis Plan

Study M16-298

A Randomized, Double-Blind, Placebo-Controlled Phase 3 Study of Rovalpituzumab Tesirine as Maintenance Therapy Following First-Line Platinum Based Chemotherapy in Subjects with Extensive Stage Small Cell Lung Cancer (MERU)

Date: 29 May 2019

Version 1.0

2.0 Table of Contents

List of Tables

List of Figures

3.0 Introduction

This statistical analysis plan (SAP) describes the detailed statistical methods for the analysis of efficacy and safety data collected in Study M16-298 (MERU) as outlined in the protocol for Study M16-298 Amendment 4 dated March 05, 2019 and will describe analysis conventions to guide the statistical programming work.

Efficacy and safety analyses will be performed using SAS Version 9.3 or higher (SAS Institute, Inc., Cary, NC 27513) under the UNIX operating system. The SAP will be finalized before the study database is unblinded.

4.0 Study Objectives and Design

4.1 Objectives

Primary Objective

- To evaluate if rovalpituzumab tesirine improves progression-free survival assessed by CRAC, and overall survival in subjects with extensive-stage SCLC with a high level of DLL3 expression ($DLL3^{high}$) who have ongoing clinical benefit (SD, PR, or CR) following the completion of 4 cycles of first line, platinum-based chemotherapy (cisplatin or carboplatin plus irinotecan or etoposide) compared to placebo.

Secondary objectives

- To evaluate if rovalpituzumab tesirine improves progression-free survival by CRAC and overall survival in all randomized subjects compared to placebo.
- To assess change in patient reported outcomes (PRO) with physical functioning as measured by the EORTC QLQ-C30 questionnaire in all randomized subjects compared to placebo.

Exploratory objectives:

- To evaluate rovalpituzumab tesirine anti-tumor activity by determining objective response rate (ORR), clinical benefit rate (CBR), and duration of responses (DOR) by CRAC and investigator assessment in DLL3^{high} and in all randomized subjects.
- To evaluate if rovalpituzumab tesirine improves progression-free survival by investigator assessment in DLL3^{high} and in all randomized subjects.
- To assess change from baseline in all patient reported outcomes (PRO) domains measured by EORTC QLQ-C-30 (except physical functioning), EORTC-QLQ-LC-13, and EQ-5D-5L in DLL3^{high} and in all randomized subjects.

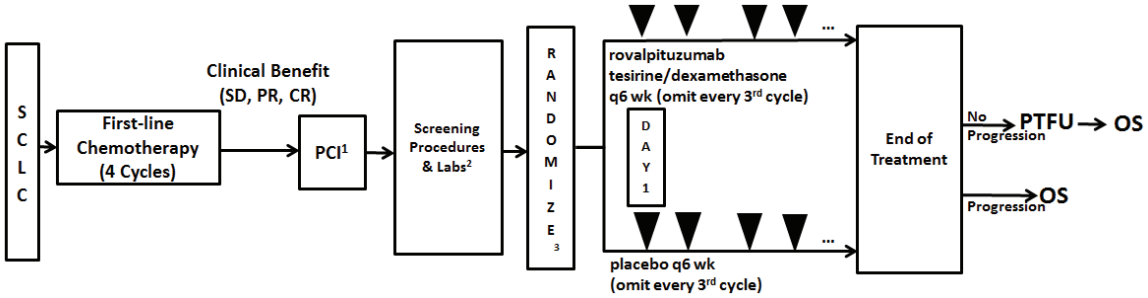
4.2 Design Diagram

This is a Phase 3, randomized, double-blinded, placebo-controlled, multinational, and multicenter study comparing the efficacy and safety of rovalpituzumab tesirine to placebo in subjects with extensive-stage SCLC with a high level of DLL3 expression (DLL3^{high}) as well as in any subjects with extensive-stage SCLC.

Eligible subjects will be randomly assigned in a 1:1 ratio to receive 0.3 mg/kg rovalpituzumab tesirine or matching placebo, and will receive their assigned therapy on Day 1 of each 6-week cycle, omitting every third cycle. Subjects will also receive 8 mg orally (PO) of dexamethasone or matching placebo twice daily on Day –1, Day 1 (the day of dosing), and Day 2 of each 6-week cycle in which rovalpituzumab tesirine/placebo is administered. Subjects will be stratified by RECIST v1.1 response after completion of first-line platinum-based chemotherapy at screening assessment (SD vs. PR/CR), DLL3 expression (Unknown vs. 0% to < 25% vs. 25% to < 75% vs. 75% or above), history of central nervous system (CNS) metastases (Yes vs. No), and for subjects with no history of CNS metastases, PCI vs. no PCI. Survival Follow-up will continue until the endpoint of death, the subject becomes lost to follow-up or withdraws consent or termination of the study by AbbVie, whichever occurs first.

A schematic of the study is provided in Figure 1, Study Schema.

Figure 1. Study Schema



Arrowheads = blinded investigational product administration; CR = complete response; OS = overall survival; PR = partial response; PTFU = post-treatment follow-up; SD = stable disease

1. Upon completion of first-line platinum-based chemotherapy, eligible subjects must be offered prophylactic cranial irradiation (PCI), if offering this procedure is not contradictory to country or institutional guidelines. Subjects receiving PCI must complete it prior to randomization into the study.
2. Collection of tumor material for DLL3 testing will be provided any time after the signing of the informed consent and prior to randomization.
3. At least 3 but no more than 9 weeks between the administration of the last cycle of platinum-based chemotherapy and randomization.

4.3 Randomization and Stratification

All subjects in the study will be randomized using an IRT system. The site will contact the IRT to obtain a Screening (subject) number once the subject has signed the informed consent. Once the screening number is assigned, if the subject is not eligible to be randomized into the study, the reason for screen failure will be documented in the source document and in the eCRF. For others, the site will access the system and a unique randomization number will be provided.

The IRT will randomize subjects in a 1:1 ratio to the rovalpituzumab tesirine treatment arm or to the placebo arm. The stratification factors used for the randomization should be the last values on or prior to the date of randomization.

The stratification factors are:

- RECIST v1.1 response after completion of first-line platinum-based therapy at screening assessment (SD vs. PR/CR),
- DLL3 expression (Unknown vs. 0% to < 25% vs. 25% to < 75% vs. 75% or above),
- History of CNS metastases (Yes vs. No),
- For subjects with no history of CNS metastases, PCI vs. no PCI.

4.4 Sample Size

There are two primary efficacy endpoints for this study: Progression-free survival (PFS) based on the CRAC and overall survival (OS) in DLL3^{high} ED SCLC subjects. To maintain the overall 1-sided type I error rate at a 2.5% for this study, the type I error will be split, assigning 0.25% to progression-free survival and the remaining 2.25% to overall survival. However, the OS endpoint will be tested as a one-sided 2.2499% significance level as a one-sided alpha of 10^{-6} will be spent for the early look at the OS data for futility analysis.

Sample size of the study is primarily determined by the analysis of OS. It is assumed, taking into account the expected patient population for the study, that median overall survival in the placebo and rovalpituzumab tesirine arm will be around 9 months and 13 months, respectively. The increase of median OS in rovalpituzumab tesirine arm corresponds to a hazard ratio of 0.69, i.e., a reduction in the hazard of death by 31%. With all these assumptions, a total of 319 deaths among subjects with DLL3^{high} are needed to achieve 90% power based on a log-rank test at a one-sided significance level of 0.022499. Assuming a 19-month accrual period and the last enrolled subject followed for 12 months, at least 480 subjects with DLL3^{high} are expected to be randomized (240 subjects in each arm). A total of 740 subjects regardless of DLL3 expression level (including 480 subjects with DLL3^{high}) will enroll in the study, assuming an approximate prevalence of 65% for subjects with DLL3^{high}.

The primary endpoints of OS and PFS will be analyzed at the same time after observing at least 319 OS events in subjects with DLL3^{high}. It is expected that approximately 420 PFS events assessed by the CRAC will be observed at the time of the primary analysis.

It is assumed, taking into account the expected patient population for the study, that median progression-free survival for the placebo and rovalpituzumab tesirine arm will be approximately 3 months and 4.5 months, respectively. The increase of median progression-free survival in rovalpituzumab tesirine arm corresponds to a hazard ratio of 0.667. With all these assumptions, a total of approximately 420 PFS events assessed by the CRAC for the subjects with DLL3^{high} are needed to achieve 91% power based on a log-rank test at a one-sided significance level of 0.0025.

4.5 Independent Data Monitoring Committee (IDMC)

An Independent Data Monitoring Committee (IDMC) will be formed and constituted according to appropriate regulatory guidelines. The IDMC will review unblinded data from the study with respect to subject accrual, pretreatment characteristics of accrued subjects, the frequency and severity of toxicities and morbidity, and provide recommendations according to the charter. Detailed information regarding the composition of the committee and procedures including frequency of meetings and scope of reviews will be documented in a separate charter.

4.6 Planned Analyses

4.6.1 Futility Analysis

A futility analysis will be conducted when approximately 160 deaths in subjects with DLL3^{high} ED SCLC (i.e., 50% of planned deaths) are observed.

The trial may be stopped for futility if the estimated Overall Survival (OS) hazard ratio of rovalpituzumab tesirine to placebo using Cox proportional-hazard regression model adjusting for randomization stratification factors exceeds 0.9. A one-sided alpha of 10^{-6} will be spent for the futility analysis for the early look at OS data.

Unblinded interim data will be analyzed and reviewed by the IDMC. The trial may be considered for early stopping for futility and the Sponsor will be unblinded to the aggregated data by randomized treatment arms if the IDMC makes such a recommendation after consideration of the OS results, and other supportive evidence including other relevant efficacy and safety data.

In the event that the IDMC recommends continuing the study due to lack of evidence for futility, the unblinded interim results will be kept confidential to the investigators, subjects, sponsors, and personnel involved in the conduct of the study until the end of the study, since the final analyses of OS and all other endpoints will be performed in an inferentially seamless manner. Subjects already enrolled in the study will continue rovalpituzumab tesirine or placebo treatment in a blinded fashion to obtain more mature OS data for a robust characterization of treatment effect.

Details for the futility analysis will be provided in the IDMC charter.

4.6.2 Final Analysis

The final analysis will be performed when at least 319 deaths have observed in DLL3^{high} subjects. It is projected that 420 PFS events assessed by CRAC would be observed in DLL3^{high} subjects at the time of the final analysis.

4.7 Type I Error Adjustment Procedures for Multiple Testing

To meet global regulatory requirements, a multiple testing strategy will be implemented to control the family-wise type I error (alpha) for comparisons of rovalpituzumab tesirine arm versus placebo arm with respect to progression-free survival (PFS) per CRAC in DLL3^{high} Set, overall survival (OS) in DLL3^{high} Set, PFS per CRAC in Randomized Set, OS in Randomized Set, and physical functioning scale score (EORTC QLQ-C30) in Randomized Set.

The following null hypotheses are considered:

H_{01a}: Rovalpituzumab tesirine arm is not superior to placebo arm in PFS per CRAC in $DLL3^{high}$ Set.

H_{01b}: Rovalpituzumab tesirine arm is not superior to placebo arm in OS in $DLL3^{high}$ Set.

H_{02}: Rovalpituzumab tesirine arm is not superior to placebo arm in OS in Randomized Set.

H_{03}: Rovalpituzumab tesirine arm is not superior to placebo arm in PFS per CRAC in Randomized Set.

H_{04}: Rovalpituzumab tesirine arm is not superior to placebo arm in physical functioning scale score (EORTC QLQ-C30) in Randomized Set.

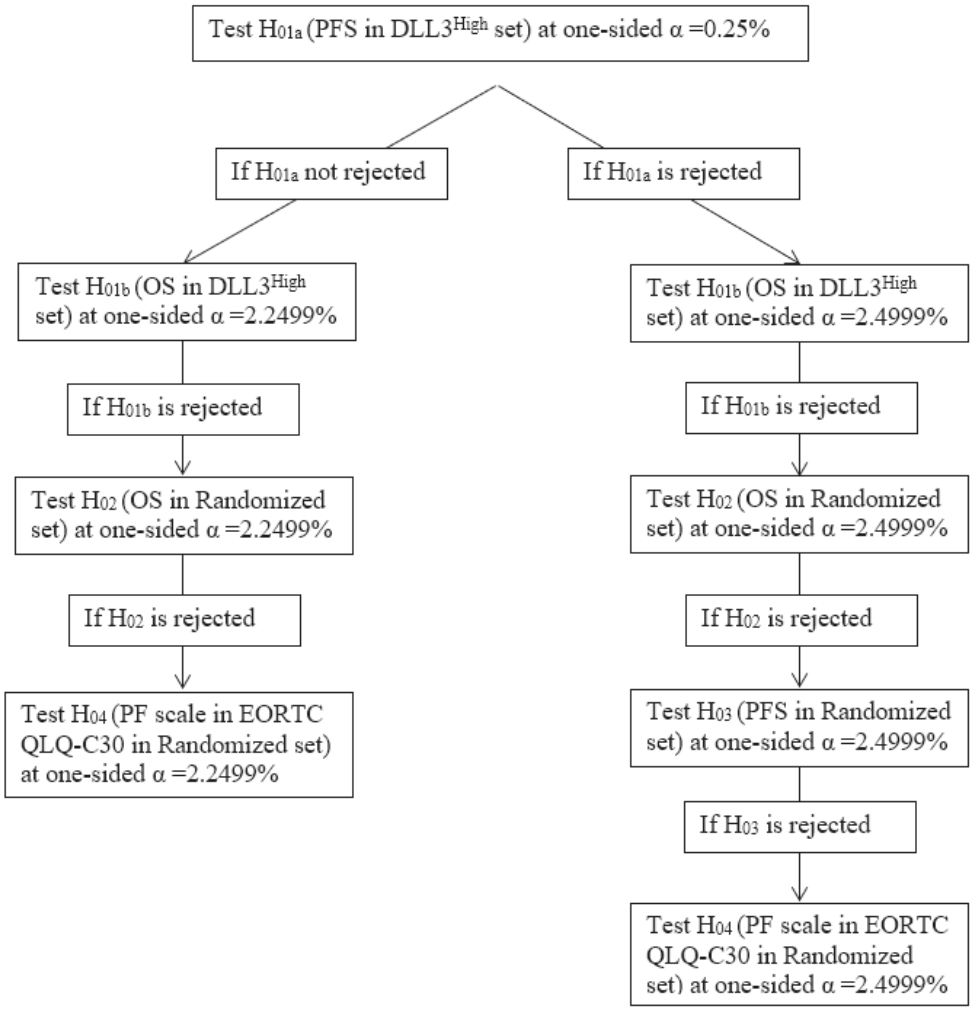
The null hypotheses will be tested in a fixed sequence of {H_{01a}, H_{01b}, H_{02}, H_{03}, and H_{04}} in order. To maintain the family-wise type I error rate for the primary endpoints at 2.5%, the null hypotheses in H_{01} (H_{01a} and H_{01b}) will be first tested with an alpha-split approach. The one-sided alpha of 0.25% and 2.25% will be assigned to PFS and OS hypotheses in H_{01}, respectively. Out of the one-sided alpha of 2.25% allocated to OS hypothesis (H_{01b}), the one-sided alpha of 10^{-6} will be spent for the early look at the OS data for futility analysis. If the hypothesis for PFS in H_{01} (H_{01a}) is rejected, the one-sided alpha of 0.25% will be recycled to the OS hypothesis in H_{01} (H_{01b}). Hence, the OS hypothesis in H_{01} (H_{01b}) will be tested either at a one-sided 2.4999% or one-sided 2.2499% alpha level of significance depending on the PFS in H_{01} (H_{01a}) is rejected or not.

If either PFS or OS endpoint reaches statistical significance, the study is considered positive. No further tests will be performed if OS hypothesis in H_{01} (H_{01b}) does not reach statistical significance.

The hypothesis in H_{02} will be tested if the hypothesis in H_{01b} is rejected. If the hypothesis in H_{02} is rejected, then, either (i) H_{03} and H_{04} will be sequentially tested if the PFS endpoint in DLL3^{high} Set (H_{01a}) shows statistical significance; or (ii) only H_{04} will be tested if the PFS endpoint in DLL3^{high} Set (H_{01a}) is not rejected.

Diagram of Hierarchical Testing of Primary and Secondary Endpoints is in Figure 2.

Figure 2. Diagram of Hierarchical Testing of Primary and Secondary Endpoints[1]



5.0 Analysis Sets

The following analysis sets will be used for analysis of safety and efficacy endpoints of the study.

DLL3^{high} Set will comprise all randomized subjects with DLL3^{high}. Subjects with unknown DLL3 status based on the sample collected at the time or before randomization be excluded. Subjects will be classified according to the treatment arm to which they are randomized regardless the actual treatment received, following intent-to-treat principle. The DLL3^{high} Set will be the primary analysis set for the analysis of efficacy endpoints of OS and PFS. All the exploratory efficacy endpoints will be analyzed based on the DLL3^{high} Set unless the OS and PFS based on the Randomized Set are statistically significant.

Randomized Set will comprise all randomized subjects, with subjects grouped according to the treatment arm to which they are randomized regardless the actual treatment received, following intent-to-treat principle. The randomized set will be the analysis set for secondary efficacy endpoints of OS and PFS, and secondary endpoint of EORTC QLQ-C30 physical functioning domain. All the exploratory efficacy endpoints will be analyzed based on the Randomized Set if the OS and PFS based on the Randomized Set are statistically significant.

Per Protocol Set (PPS) will comprise all subjects in the randomized set without any major protocol violations which may affect the evaluation of the primary efficacy endpoint.

Subjects meeting any of the following criteria will be excluded or censored from PPS:

- Those entered into the study even though did not satisfy the entry criteria
- Those who developed withdrawal criteria during the study but were not withdrawn will be censored when they develop the withdrawal criteria
- Those who received the wrong treatment
- Those who received an excluded concomitant treatment
- Those who were randomized but did not receive any dose of study medication.

Subjects will be classified according to treatment assigned at the time of randomization. The Per Protocol Set will be used for supportive analysis of efficacy endpoints.

Safety Set will comprise all subjects who receive at least one dose of study drug and subjects will be classified according to treatment received. Thus a subject who is randomized to the rovalpituzumab tesirine arm but does not receive rovalpituzumab tesirine will be considered in the placebo arm for safety analysis. Dexamethasone treatment will not be taken into consideration for subjects' classification in the Safety set.

6.0 Analysis Conventions

This section provides general considerations for data handling, summary, and analysis.

Continuous variables will be summarized by sample size (N), mean, standard deviation (St. Dev.), median, minimum, and maximum. Frequency and percentage will be provided for categorical variables and 95% confidence intervals (CIs) will be generated for parameter estimates of interest, unless specified otherwise.

6.1 Dealing with Multiple Values on the Same Day

In cases where multiple values are collected on the same day at baseline or post-baseline visit, either the arithmetic average (e.g., for continuous summary) or worst value (e.g., for shift tables) will be used for further analyses.

6.2 Definition of Baseline

Unless otherwise specified, the baseline observation is defined as the last non-missing measurement collected prior to or on the first dose date of blinded investigational product (rovalpituzumab tesirine or matching placebo; dexamethasone or matching placebo).

6.3 Definition of Final Observation

For lab and vital sign parameters, the final observation is defined as the last non-missing post-baseline measurement collected not more than 70 days after the last dose of study drug.

6.4 Missing Data

No missing data imputation will be implemented, unless specified otherwise.

6.5 Definition of Study Days and Rx Days

Study Day of any observation is defined for post-randomization observations as:

Study Day = Date of observation – Date of randomization + 1,

and for observations pre-dating randomization as:

Study Day = Date of observation – Date of randomization.

Thus, the day of randomization is defined as Study Day 1, while the day prior to the randomization is defined as Study Day –1 (there is no Study Day 0).

Study Rx Day of any post-baseline observation is defined as the number of days from the day of the first dose of any study drug (rovalpituzumab tesirine, Dexamethasone, matching placebo) to the date of observation. It is calculated for each post-treatment observation as follows:

Study Rx Day = Date of observation – Date of first dose of any study drug + 1

Definition of Analysis Windows

During the treatment period, all time points and corresponding time windows are based on Study Rx Days, unless specified otherwise. For longitudinal analyses, especially for the laboratory data (hematology, chemistry, and urinalysis), the time windows specified below describe how data collected at protocol specified visits will be assigned to. If more than one observation is included in a time window, the observation closest to the nominal day should be used. If there are two observations equally distant to the nominal day, the later one will be used in analyses.

Table 1. Visit Window for Laboratory[#], Vital Sign Assessments and Physical Examination

Study Visit (Week)[*]	Nominal Study Rx Day	Time Window (Study Rx Days Range)
Baseline	1	≤ 1
3	22	2 to ≤ 32
6	43	33 to ≤ 53
9	64	54 to ≤ 74
12	85	75 to ≤ 106
18	127	107 to ≤ 137
21	148	138 to ≤ 158
24	169	159 to ≤ 179
27	190	180 to ≤ 200
30	211	201 to ≤ 232
…	…	…
…	…	…
18•X	7•18•X + 1	7•18•X - 19 to ≤ 7•18•X + 11
18•X + 3	7• (18•X + 3) + 1	7•18•X + 12 to ≤ 7•(18•X + 3) + 11
18•X + 6	7• (18•X + 6) + 1	7•(18•X + 3) + 12 to ≤ 7•(18•X + 6) + 11
18•X + 9	7• (18•X + 9) + 1	7•(18•X +6) + 12 to ≤ 7•(18•X + 9) + 11
18•X + 12	7• (18•X + 12) + 1	7•(18•X +9) + 12 to ≤ 7•(18•X + 12) + 22
…	…	…

laboratory tests include hematology, serum chemistry and urinalysis.

* Every third cycle (6-week cycle) treatment will be omitted.

Table 2. Visit Window for Coagulations Tests and Patient Reported Outcomes (PRO)

Study Visit (Week)*	Nominal Study Rx Day	Time Window (Study Rx Days Range)
Baseline	1	≤ 1
6	43	2 to ≤ 64
12	85	65 to ≤ 106
18	127	107 to ≤ 148
24	169	149 to ≤ 190
30	211	191 to ≤ 232
…	…	…
…	…	…
6•X	7•6•X + 1	7•6•X + 1 - 20 to ≤ 7•6•X + 1 + 21
…	…	…

* Every third cycle (6-week cycle) treatment will be omitted.
Coagulation tests are performed at screening visit, first day of each cycle, and end of treatment visit.

7.0 Demographics, Baseline Characteristics, Medical History, Previous/Concomitant Medications, and Prior Oncology Therapies

Data for demographic, baseline characteristics, medical/surgical history, previous/concomitant medications and prior oncological procedures will be summarized for each treatment arm and overall on both the DLL3^{high} Set and Randomized Set.

7.1 Demographic and Baseline Characteristics

The following demographic, current disease history and baseline disease characteristics (including stratification variables) will be summarized by treatment groups.

Demographic variable	Baseline disease characteristics
• Age • Weight • Sex • Race (White, Black or African American, Asian, American Indian or Alaskan Native, Native Hawaiian or Pacific Islander) • Ethnicity (Hispanic or Latino, Not Hispanic or Latino, Other) • Age ○ < 40, 40 - < 65, ≥ 65 yr • History of tobacco product use and alcohol use (current, former, never, unknown) • Region of the world (North America, Asia, Europe, Rest of the world)	• TNM staging at the start of the FL treatment • RECIST v1.1 response after completion of first line platinum-based therapy at screening assessment (SD, CR/PR) • History of CNS metastases (Yes, No) • PCI status (PCI, No PCI, NA) • DLL3 expression level (unknown, < 25%, 25% - < 75%, ≥ 75%) • ECOG performance score • LDH: ≤ ULN vs. > ULN
	Baseline disease status
	• Time from initial diagnosis to start of randomization • Time from most recent prior cancer therapy (as defined by C4D1 in the frontline platinum therapy) to randomization • Number of metastatic sites, 0 – 1 vs. > 1

Continuous data (e.g., age, height, and weight) will be summarized with N, mean, standard deviation (St. Dev.), median, minimum, and maximum. Frequencies and percentages will be computed for the categorical variables.

There will be no statistical comparison for demographic and baseline measurements.

7.2 Medical History

Medical history data will be summarized and presented using conditions/diagnoses as captured on the eCRF. The conditions/diagnoses will be presented in alphabetical order. The number and percentage of subjects with a particular condition/diagnosis will be summarized for each treatment arm. Subjects reporting more than one condition/diagnosis will be counted only once for that condition/diagnosis. There will be no statistical comparison for the medical history between treatment arms.

7.3 Prior and Concomitant Medications

A prior medication is defined as any medication with start date prior to the date of randomization and collected on the electronic case report forms (eCRFs). A concomitant medication is defined as any drug that taken on the day of or after randomization, but not 70 days after the last study drug administration (rovalpituzumab tesirine or matching placebo; dexamethasone or matching placebo) regardless of the start and stop date. Medications taken on the day of the first dose of study drug are counted as concomitant medications.

For reporting purpose, the following conservative approach will be followed for determination of prior and concomitant medication:

Start date	**End date**	
Prior to Study Day 1	Prior to Study Day 1	Prior medication
Missing	Prior to Study Day 1	Prior medication
Prior to Study Day 1	Study Day 1 to Last dose + 70 day	Both prior medication and concomitant medication
Study Day 1 to Last dose + 70 day	Study Day 1 to Last dose + 70 day	Concomitant medication
Missing	Study Day 1 to Last dose + 70 day	Concomitant medication
Study Day 1 to Last dose + 70 day	Missing	Concomitant medication
Missing	Missing	Concomitant medication

The frequency and percentage of subjects who took at least one dose of medication other than study drug (rovalpituzumab tesirine and dexamethasone) will be summarized by the anatomical therapeutic chemical (ATC) codes in the World Health Organization (WHO) Drug dictionary. Prior and concomitant medications will be summarized separately.

A similar summary will be prepared for prior medications.

There will be no statistical comparison for the prior and concomitant medications between treatment arms.

7.4 Prior Anti-Cancer Therapy and Surgery

For each prior anti-cancer therapy/surgery, the frequency and percentage of subjects who had the prior anti-cancer therapy/surgery will be provided. In addition, the number of cycles and treatment duration will be summarized for prior anti-cancer therapy.

There will be no statistical comparison for prior anti-cancer treatment/surgery between treatment arms.

8.0 Subject Disposition

Subject disposition (number and percentage of subjects) will be presented on all subjects who are screened by investigator site and overall. The following information will be presented:

- Subjects who were screened
- Subjects who screen failed
- Subjects who were randomized
- Randomized subjects who were DLL3^{high}
- Subjects who were included in the Per Protocol Set
- Subjects who took at least 1 dose of blinded study drug
- Subjects who discontinued the study drug (by primary reason and by any reason)
- Subjects who discontinued the study (by primary reason and by any reason)

Primary and all reasons for study drug discontinuation will be summarized by each of rovalpituzumab tesirine and dexamethasone.

No statistical tests will be performed.

9.0 Study Drug Exposure

Analyses for rovalpituzumab tesirine exposure will be performed on the Safety Set.

Duration of rovalpituzumab tesirine treatment, defined as the total number of weeks a subject received rovalpituzumab tesirine, accounting for (1) study drug was administered on Day 1 of each 6-week cycle, and (2) every 3rd cycle was omitted:

Treatment cycle #*	Duration of Rova-T treatment (weeks)	Theoretical duration of Rova-T treatment (weeks)	Planned dose (mg/kg)
1 + 3 × X†	(Date of last dose – date of first dose + 42)/7	6 × (1 + 3 × X)	0.3 + 0.6 × X
2 + 3 × X	(Date of last dose – date of first dose + 84)/7	6 × 3 × (1 + X)	0.6 × (1 + X)

* The last treatment cycle that subject received non-zero dose of Rova-T.

† X = 0, 1, 2, … corresponding to Cycle 1, 2, 4, 5….

Summary statistics for duration of rovalpituzumab tesirine treatment will be presented by treatment arm for the following categories: < = 6 weeks, 6 to < = 18 weeks, 18 to ≤ 24 weeks, 24 to ≤ 36 weeks, etc.

Cumulative dose, defined as the sum of the doses (mg/kg) administered to a subject during the treatment period, and will be summarized by N, mean, standard deviation (St. Dev.), median, minimum, and maximum. Summary statistics for average dose administered by treatment arm will be presented at each cycle.

Dose intensity (mg/kg/week), defined as the ratio of the total dose received to the total duration of exposure.

$$DI_{observed} = \frac{Cumulative\ dose\ (mg/kg)}{Duration\ of\ treatment\ (weeks)}$$

Relative dose intensity 1 (RDI_1) is calculated as the ratio of observed dose intensity to the theoretical dose intensity indicated in the protocol, expressed in percent (%).

$$DI_{protocol} = \frac{Planned\ dose\ \left(\frac{mg}{kg}\right)}{Theoretical\ duration\ of\ treatment\ (weeks)}$$

$$RDI_1(\%) = \frac{DI_{observed}}{DI_{protocol}} \times 100$$

Relative dose intensity 2 (RDI_2) is calculated as the ratio of cumulative dose to the planned dose indicated in the protocol, expressed in percent (%).

$$RDI_2(\%) = \frac{Cumulative\ dose\ (mg/kg)}{Planned\ dose\ (mg/kg)} \times 100$$

Summary statistics for RDI_1 and RDI_2 will be presented. The number of subjects in the following categories (e.g., RDI < 70%, 70% to < 90%, 90% to < 110%, 110% to < 120%, 120% or higher) will also be presented.

10.0 Efficacy Analysis

10.1 General Considerations

Efficacy endpoints of response and progression used in analyses per Central Radiographic Assessment Committee (CRAC) and investigator assessment will be based RECIST v1.1. For subjects whose tumor meets the criteria of complete or partial response (CR or PR), tumor response must be confirmed by repeated assessments. Disease progression will be defined as radiographic progression of disease.

The statistical comparisons for the primary and secondary endpoints will be done according to the multiplicity adjustment plan described in Section 4.7 to preserve the familywise one-sided type 1 error rate to 2.5%. The statistical comparisons for all the exploratory efficacy endpoints will be performed at a nominal one-sided 2.5% significance level. Confidence intervals for parameters (e.g., median survival times, HR) will be constructed with 2-sided 95% confidence level, whenever applicable. No type I error adjustment for multiple comparisons will be carried out for these exploratory endpoints.

10.1.1 Statistical Analysis Methods

Time-to-Event Endpoints

All time to event (TTE) endpoints are defined using a time variable and an event indicator. All TTE efficacy endpoints in this study are concerned with only the

first occurrence of the event of interest, all recurrence of the same event is not considered for endpoint derivation or analysis. An event however may be defined in a composite manner, i.e., as the occurrence of one among several different outcomes. The composite event is observed when at least one of the component events occurs, and the time to the earliest among the occurring component events is considered to be the TTE for the composite event.

The time variable will be computed in days and converted into months (1 month = 30.4375 days). Thus, for subjects who experience the event of interest, time to event will be defined as the time from randomization to the first occurrence of the event. When multiple assessments are needed to ascertain the occurrence of an event, the earliest date among all of these assessments is taken to be date of the event or censoring. For a subject who does not experience an event on the study, time will be right censored at the time of his last available adequate post-baseline assessment that rules out the occurrence of the event. If a subject had no post-baseline assessment and did not experience the event of interest, then TTE will be censored at the date of randomization and we will set TTE = 1 day, by convention.

Follow-up time for TTE endpoints will be derived as follows:

TTE = Date of first occurrence of an event or censoring – date of randomization + 1.

TTE endpoints will be summarized by number of events observed, number of subjects censored, and times corresponding to 25% (1st quartile), 50% (median) and 75% (3rd quartile) event probabilities. Kaplan-Meier (KM) estimates will be calculated and plotted. Comparison between treatment arms will be based on stratified log-rank tests as specified.

A Cox proportional-hazards regression model with the treatment group and stratification factors as covariates will be used to estimate the hazard ratio (HR) of rovalpituzumab tesirine to placebo and its two-sided 95% confidence interval (CI).[2,3] Only the main effects of the stratification variables will be fitted.

In the presence of sparse strata that could potentially lead to biased treatment effect estimation and/or model convergence issue, the following modified stratification factors would be used:

- RECIST v1.1 response after completion of first-line platinum-based therapy at screening assessment (SD vs. PR/CR),
- DLL3 expression (Unknown or 0% to < 75% vs. 75% or above),
- History of CNS metastases (Yes vs. No),
- For subjects with no history of CNS metastases, PCI vs. no PCI.

In cases the treatment effect estimation and/or model convergence issues persist with above modified stratification factors, and RECIST v1.1 response after completion of first-line platinum-based therapy has been confirmed not prognostic of disease outcome, this factor could be removed from both stratified log-rank test and Cox proportional hazards model.

Measure of treatment effects will be provided in terms of HR (rovalpituzumab tesirine to Topotecan) from the above model with HR value less than 1 representing treatment benefit.

Continuous Variables and Changes from Baseline

Summary of continuous endpoints at baseline, post-baseline and change from baseline (as available) will be provided by treatment arms. Change from baseline values will be analyzed using an analysis of covariance (ANCOVA) model including stratification factors and baseline as covariates and 95% CI for treatment group difference will be provided. For change from baseline analysis of PROs, patients with missing baseline value will be excluded. These summaries will also be further broken down by individual levels of each stratification factors.

Categorical Variables and Response Rates in the Study

Summary of categorical endpoints (e.g., frequency, percentage) will be provided by treatment arms. The response rates in treatment arms will be compared using Cochran-Mantel-Haenszel test stratified by randomization stratification factors. Rates and 95% confidence intervals will be reported for each treatment arm.

Stratification Factors Used in Statistical Tests and Models

Since approximately 75% of subjects enrolled are expected to have DLL3 expression ≥ 75%, very sparse data is expected in the three remaining strata for DLL3 expression (namely, Unknown, 0% to < 25% and 25% to < 75%) that can lead to loss of power for the stratified log rank test, Cochran-Mantel-Haenszel test, ANCOVA, and result in instability and biases in estimation of parameters of the Cox regression model.

Hence, all categories of DLL3 expression < 75% or unknown will be combined into one category for analysis. If a statistical estimation procedure does not converge even after simplification of the DLL3 expression-based stratification, additional variable selection methods will be employed based on univariate assessment of the prognostic importance of individual factors on the associated endpoint (using an ad hoc p-value threshold of 0.2) and consideration of pairwise correlations between factors. If statistical models, tests and estimation remain unstable after the previous step or no other factors can be dropped from analysis, an unstratified analysis will be performed.

10.2 Efficacy Endpoints and Analyses

10.2.1 Primary Efficacy Endpoints

The primary efficacy endpoints are:

- Progression-free survival (PFS) determined by the CRAC
- Overall survival (OS)

Progression Free Survival (PFS)

PFS will be defined as the number of months from randomization to disease progression, as assessed by the CRAC per RECIST v1.1, or death of any cause, whichever occurs first. Any death that occurs > 12 weeks (two consecutive missing scans) after the previous radiographic disease assessment, or any PD or death happened after the initiation of new anti-cancer therapy will be excluded disregarded from the analysis. Subject without an event will be censored at the last date of post-baseline radiographic assessment. Subject with no post-baseline assessment and did not experience death event within 12 weeks of randomization will be censored at the date of randomization.

Differences between the treatment arm and the placebo arm in PFS will be assessed by a log-rank test stratified by the randomization stratification factors, testing the null hypothesis (rovalpituzumab tesirine arm is not superior to placebo arm in PFS). The hazard ratio of the treatment arm over the placebo arm will be calculated by a Cox's proportional hazards regression model adjusting for the main-effects of the stratification factors. Median PFS and 95% confidence intervals will be estimated using Kaplan-Meier survival methodology, with the Kaplan-Meier survival curves presented to provide a visual description. PFS rates at 3, 6, 9, and 12 months after randomization with 95% confidence intervals will be reported. Estimates of the treatment effect will be expressed as hazard ratio including 95% confidence intervals estimated through a Cox proportional-hazards regression model adjusting for the main-effects of the stratification factors. The same methodology will be applied to PFS based on investigator assessment.

Overall Survival (OS)

OS is defined as time from the randomization to death due to any cause. For subjects who are alive at the end of this study, OS will be right-censored on the last date the subject is known to be alive.

Differences between the treatment arm and the placebo arm in OS will be assessed by a log-rank test stratified by the randomization stratification factors, testing the null

hypothesis (rovalpituzumab tesirine arm is not superior to placebo arm in OS). The hazard ratios of the treatment arm over the placebo arm will be calculated by a Cox's proportional hazards regression model adjusting for the main-effects of the stratification factors. Median OS and 95% confidence intervals will be estimated using Kaplan-Meier survival methodology, with the Kaplan-Meier survival curves presented to provide a visual description. OS rate at 6, 12, 18, and 24 months and corresponding 95% confidence intervals will be reported. Estimates of the treatment effect on will be expressed as hazard ratio including 95% confidence intervals estimated through Cox proportional-hazards regression models adjusting for the main-effects of the stratification factors.

10.2.2 Secondary Efficacy Endpoints

The secondary efficacy endpoints in Randomized Set are:

- OS
- PFS determined by the CRAC
- EORTC QLQ-C30 physical functioning domain

OS in the Randomized Set and PFS by CRAC in Randomized Set will be analyzed using similar statistical methods described for the primary endpoints.

Change from baseline of the items and domains of the EORTC QLQ-C30 physical functioning domain will be summarized by treatment arm. The treatment group differences will be evaluated by analyzing the change from baseline using the analysis of covariance (ANCOVA) model. Detailed scoring methods are provided in Section 10.2.3.

10.2.3 Exploratory Efficacy Endpoints

The exploratory efficacy endpoints in both DLL3^{high} Set and Randomized Set are:

- Objective Response Rate (ORR) per CRAC
- PFS per investigator assessment
- ORR per investigator assessment

- Clinical Benefit Rate (CBR) per CRAC and investigator assessment, respectively
- Duration of Response (DOR) per CRAC and investigator assessment, respectively
- Change from baseline in all PRO domains (except physical functioning) measured by EORTC QLQ-C30/LC13 and EQ-5D-5L.

Objective Response Rate (ORR)

ORR is defined as the proportion of patients with a best overall response (BOR) of confirmed complete response (CR) or confirmed partial response (PR) by CRAC review according to RECIST v1.1.

All subjects with measurable disease who were randomized regardless of whether they had any post-baseline assessment will be included in the analysis. Any subject who did not meet confirmed CR or PR criteria, including those who did not have post-baseline radiological assessments will be considered as non-responder.

ORR in treatment arms will be compared using Cochran-Mantel-Haenszel test stratified by randomization stratification factors. Rates and 95% confidence intervals will be reported for each treatment arm.

Similarly, analysis of ORR assessed by investigator will be performed.

Clinical Benefit Rate (CBR)

CBR is defined as the proportion of patients with a BOR of CR or PR, or stable disease (SD) by CRAC review according to RECIST v1.1.

All subjects with measurable disease who were randomized regardless of whether they had any post-baseline assessment will be included in the analysis. Any subject who did not meet CR, PR, or SD criteria, including those who did not have post-baseline radiological assessments will be considered as having no clinical benefit.

The same methodology as for the analysis of ORR will be applied. Similarly, analysis of CBR assessed by investigator will be performed as well.

Similarly, analysis of CBR assessed by investigator will be performed.

Duration of Response (DOR)

The duration of response (DOR) for a given subject will be defined as the number of months from the day the criteria are met for confirmed complete response (CR) or partial response (PR) by CRAC (whichever is recorded first) to the date of progressive disease or death, whichever occurs first. If a subject is still responding (i.e., has not progressed nor died after CR or PR), then the subject's data will be censored at date of the last radiographic assessment by CRAC. For subjects who never experienced CR or PR, these subjects' data will not be included in the DOR analysis. Distribution of the DOR will be estimated for each treatment arm using Kaplan-Meier methodology. Median DOR with corresponding 95% CI for each treatment arm will be provided.

The same methodology will be applied to analysis of the duration of response per investigator assessment.

Patients Reported Outcomes (PRO)

PRO assessments are performed at screening, Cycle 2 Day 1 only, the first Non-Treatment cycle, end of treatment (EOT), and post-treatment follow-up visits (if applicable). The EORTC QLQ-C30/LC13 scoring manual will be used to transform the raw scores into the domain scores (global heath, functional scales, symptom scales/items). Change from baseline of the items and domains of the EORTC QLQ-C30/LC13 will be summarized by treatment arm. The EQ-5D-5L manual and the published weights will be used to convert the individual items to the utility scores. Change from baseline of the EQ-5D-5L utility score and VAS will be summarized by treatment arm. The treatment group differences will be evaluated by analyzing the change from baseline using the analysis of covariance (ANCOVA) model.

The EORTC QLQ-C30 is composed of global health status/QoL scale; five functional scales (physical, role, emotional, cognitive, and social); three symptom scales (fatigue, nausea and vomiting, and pain); and six single items (dyspnea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties).[4]

EORTC QLQ-C30	Item (Question) Numbers	Item Range*
Global health status/QoL		
Global health status/QoL	29, 30	6
Functional scales		
Physical functioning	1, 2, 3, 4, 5	3
Role functioning	6, 7	3
Emotional functioning	21, 22, 23, 24	3
Cognitive functioning	20, 25	3
Social functioning	26, 27	3
Symptom scales/Items		
Fatigue	10, 12, 18	3
Nausea and vomiting	14, 15	3
Pain	9, 19	3
Dyspnea	8	3
Insomnia	11	3
Appetite loss	13	3
Constipation	16	3
Diarrhea	17	3
Financial difficulties	28	

* Item range is the difference between the possible maximum and the minimum responses to individual items; most items take values from 1 to 4, giving range = 3.

The lung cancer module (QLQ-LC13) is meant for use among lung cancer subjects in parallel with the core QLQ-C30. It is composed of one multi-item scale to assess dyspnea, and a serious of single items assessing coughing, sore mouth, dysphagia, peripheral neuropathy, alopecia, hemoptysis, and pain.

QLQ-LC13	Item (Question) Numbers	Item Range*
Symptom scales/items		
Dyspnea	3, 4, 5	3
Coughing	1	3
Hemoptysis	2	3
Sore mouth	6	3
Dysphagia	7	3
Peripheral neuropathy	8	3
Alopecia	9	3
Pain in chest	10	3
Pain in arm or shoulder	11	3
Pain in other parts	12	3
Pain medication	13	3

* Item range is the difference between the possible maximum and the minimum responses to individual items.

All scales and single items are reported on categorical responses and will be linearly converted to 0 through 100 numeric values. Given a scale, the raw score (RS) is computed as the mean of component items over the number of items answered for that scale. Then the scale score (SS) will be computed as following.

For Global health status/QoL, and Symptom scales/Items,

$$SS = \{(RS - 1)/range\} \bullet 100.$$

For Functional scales,

$$SS = \{1 - (RS - 1)/range\} \bullet 100.$$

If 50% or more of the items are answered for a given scale, the scale score will be computed as described above. If not, the scale score will not be computed. For example, physical functioning contains 5 items (questions 1 through 5), and the scale score is calculated if at least 3 of the items are answered. For single-item measures (dyspnea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties), none of the

single-item measures can be computed if not answered. If a scale score cannot be computed, the outcome for that score is left blank.

The scoring approach for the QLQ-LC13 is identical in principle to that for the symptom scales/single items of the EORTC QLQ-C30.

For each scale, the treatment group differences will be evaluated by analyzing the change from baseline each of the post-baseline assessments and to the final measurement (last PRO assessment per subject) using the analysis of covariance (ANCOVA) model. Simple summary statistics for QoL assessments at each timepoint will also be presented.

Changes in QoL scores at the scale level or at the domain level per time point will be evaluated by classifying them according to the 10-point change threshold into 3 categories:[5]

- Improved is defined as a 10 point or more improvement (i.e., increase for functional scales and decrease for symptom scales) from baseline.
- Stable is defined as a less than 10 point change from baseline.
- Deteriorated is defined as a 10 point or more deterioration (i.e., decrease for functional scales and increase for symptom scales) from baseline. Subjects without a valid QoL outcome will be considered as having deteriorated.

The EuroQol 5 Dimensions 5 Level (EQ-5D-5L) is a preference based measure of health status that consists of EQ-5D descriptive system and EQ visual analogue scale (VAS).

The EQ-5D descriptive system comprises the following 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each of the dimensions is divided into 5 levels of perceived problems: Level 1 (indicating no problem), Level 2 (indicating slight problems), Level 3 (indicating moderate problems), Level 4 (indicating severe problems), and Level 5 (indicating extreme problems). EQ-5D-5L health status, defined by the EQ-5D-5L descriptive system, will be converted into a single preference – weighted health status or "utility" index score by applying country-specific weights (if

available) or U.S. weights (if not available). If country is not available in the look-up table, then weights based on the United States will be applied.

The EQ VAS records the respondent's self-rated health on a visual analogue scale (0 to 100). The VAS score will be measured separately.

The analyses will be performed in the DLL3^{high} and randomized sets. Mean change from baseline in EQ-5D-5L "utility" index score to each applicable post-baseline time point, and to the final measurement will be presented by treatment arm. The treatment group difference will be evaluated by analyzing the change from baseline using the analysis of covariance (ANCOVA) model. The same methodology will apply to the mean change from baseline in VAS score.

10.3 Sensitivity Analyses

Efficacy analyses for the primary endpoints (PFS per CRAC and OS) will be performed in the Per Protocol Set.

Two sensitivity analyses will be carried out for PFS endpoint by implementing different censoring rules in the definition of PFS (i.e., modified PFS endpoint). These sensitivity analyses will be conducted in either DLL3^{high} Set or Randomized Set.

Sensitivity analysis 1: PFS will be defined as the number of months from randomization to disease progression, as assessed by the CRAC per RECIST v1.1, or death of any cause, whichever occurs first. Subjects without an event will be right-censored at the date of last follow up for disease progression. Patients with no post baseline follow up for progression and who doesn't experience a death event will be censored at the day of randomization.

Sensitivity analysis 2: PFS will be defined as the number of months from randomization to disease progression, as assessed by the CRAC per RECIST v1.1, or death of any cause, whichever occurs first. Subjects with 2 or more consecutive missed assessments (i.e., > 12 weeks since last scan) without any subsequent assessment prior to the PFS

event (i.e., progression or death) will be considered progressed at the planned date of the earliest missed scan date (i.e., last scan date prior to event + 6 weeks). Subjects without an event or receives new anti-cancer therapy will be right-censored at the date of last follow up for disease progression.

10.4 Subgroup Analyses

In the DLL3^{high} or randomized sets, treatment effects for the primary endpoints (PFS per CRAC and OS) will be assessed in the following subgroups:

- RECIST v1.1 response after completion of first-line platinum-based therapy at screening assessment (SD vs. PR/CR)
- DLL3 expression (Unknown vs. 0% to < 25% vs. 25% to < 75% vs. 75% or above) for the randomized set only
- History of CNS metastases (Yes vs. No)
- For subjects with no history of CNS metastases, PCI vs. No PCI
- Lactate dehydrogenase (LDH), ≤ ULN vs. > ULN
- Number of metastatic sites, 0 – 1 vs. > 1
- Age group (age < 40 vs. 40 to < 65 vs. 65 or older)
- Gender
- Race
- ECOG (0 vs. 1)

The relationship between DLL3 expression level and outcome measures of OS and PFS by CRAC will be explored using additional subgroup analysis, as well as by multivariable Cox's model.

Differences between rovalpituzumab tesirine arm and the placebo arm for time-to-event endpoints will be assessed by an unstratified log-rank test for each subgroup. Hazard ratios of treatment arm over the placebo arm will be calculated by a Cox's model.

11.0 Safety Analysis

11.1 General Considerations

Unless otherwise specified, safety analyses will be performed on the Safety Set. No statistical comparisons will be performed between arms. No p-values and CIs will be provided.

Unless specified, all summaries/analyses involving AEs will only include treatment-emergent adverse events (TEAEs). TEAEs are defined as any adverse event with onset or increase severity after the first dose of study drug (rovalpituzumab tesirine or matching placebo; dexamethasone or matching placebo) and no more than 70 days after the last dose of study drug administration. AEs where the onset date is the same as the study drug start date are assumed to be treatment-emergent, unless the study drug start time and the AE start time are collected and the AE start time is prior to the study drug start time. If an incomplete onset date was collected for an AE, the AE will be assumed to be treatment-emergent unless there is evidence that confirms that the AE was not treatment-emergent (e.g., the AE end date was prior to the date of the first dose of study drug).

Adverse events will be coded and summarized by system organ class and preferred term (and severity) according to the Medical Dictionary for Regulatory Activities (MedDRA) version 21.0 or higher. The table of clinical toxicity grades modified from the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE) version 4.0 will be used in the grading of adverse events and laboratory abnormalities that are reported as adverse events.

Summary of continuous safety endpoints (e.g., change from baseline values in laboratory values, vital signs parameters) will include the mean, standard deviation, median and range. In the context of change from baseline analyses, summary of baseline and post-baseline data will also be provided. Categorical safety endpoints (e.g., incidence of AEs or clinically significant ECG values) will be summarized using frequencies and percentages. All the analyses will be carried out by treatment group.

Unscheduled assessment will not be included in the summary of change from baseline, but will be included in producing shift tables and summary of lab abnormalities.

11.2 Analysis of Adverse Events

The number and percentage of subjects having treatment-emergent adverse events will be tabulated by MedDRA system organ class (SOC) and preferred term (PT) for each treatment group. Subjects reporting more than one adverse event for a given MedDRA PT will be counted only once for that term. Subjects reporting more than one type of adverse event within a MedDRA SOC will be counted only once for that SOC. Subjects reporting more than one type of adverse event will be counted only once in the overall total.

The tabulations will also be provided with further breakdowns by maximum severity per NCI CTCAE v4.0 grade. If a subject has an AE with unknown severity, then the subject will be counted in the severity category of "unknown," unless the subject does not have another occurrence of the same AE with a severity present. If a subject has multiple occurrences of an AE with the same organ class and preferred term, the most severe event will be presented.

Adverse events possibly related to study drug (rovalpituzumab tesirine or dexamethasone) as assessed by the investigator in each treatment arm. If a subject has an AE with unknown relationship, then the subject will be counted in the relationship category of "unknown," unless the subject does not have another occurrence of the same AE with a relationship present.

Serious adverse events, adverse events leading to dose reduction, discontinuation of treatment, and adverse events leading to death will be summarized.

An overview of AEs will be presented, by treatment arm and overall, consisting of the number and percentage of subjects experiencing the following adverse event categories:

- Any treatment-emergent adverse event

- Any treatment-emergent adverse event NCI toxicity grade 3 or 4
- Any treatment-emergent adverse event that is rated as a reasonable possibility of being related to study drug
- Any treatment-emergent adverse event NCI toxicity grade 3 or 4 that is rated as a reasonable possibility of being related to study drug
- Any treatment-emergent serious adverse event
- Any treatment-emergent serious adverse event NCI toxicity grade 3 or 4
- Any treatment-emergent serious adverse event that is rated as a reasonable possibility of being related to study drug
- Any treatment-emergent adverse event leading to discontinuation of rovalpituzumab tesirine
- Any treatment-emergent adverse event leading to delay/interruption of rovalpituzumab tesirine
- Any treatment-emergent adverse event leading to reduction of rovalpituzumab tesirine
- Any treatment-emergent adverse event leading to death
- Any treatment-emergent adverse event leading to death that is rated as a reasonable possibility of being related to study drug
- Any treatment-emergent serious adverse event leading to discontinuation of rovalpituzumab tesirine
- Any treatment-emergent serious adverse event leading to delay/interruption of rovalpituzumab tesirine
- Any treatment-emergent serious adverse event leading to reduction of rovalpituzumab tesirine
- Any treatment-emergent serious adverse event leading to death

11.3 Adverse Events of Special Interest (AESIs)

Treatment-emergent AESIs, including important identified risks and potential risks, will be identified by the search criteria provided in Table 3.

Table 3. Adverse Events of Special Interest

Adverse Events of Special Interest	MedDRA Search Criteria/Codes
important identified risks	
Pleural Effusion	PTs: 10035598 Pleural Effusion 10003522 Aspiration Pleural Cavity 10063045 Effusion
Pericardial Effusion	PTs: 10034474 Pericardial Effusion 10007610 Cardiac Tamponade 10034471 Pericardial Drainage 10063045 Effusion
Generalized Edema	PT 10018092 Generalised Oedema
Photosensitivity Reaction	PT 10034972 Photosensitivity Reaction
Important Potential Risks	
Pneumonitis	PTs: 10035742 Pneumonitis 10035745 Pneumonitis Chemical 10037765 Radiation Pneumonitis 10066728 Acute Interstitial Pneumonitis 10022611 Interstitial Lung Disease

Table 3. Adverse Events of Special Interest (Continued)

Adverse Events of Special Interest	MedDRA Search Criteria/Codes
important identified risks	
Pleural Effusion	PTs: 10035598 Pleural Effusion 10003522 Aspiration Pleural Cavity 10063045 Effusion
Pericardial Effusion	PTs: 10034474 Pericardial Effusion 10007610 Cardiac Tamponade 10034471 Pericardial Drainage 10063045 Effusion
Generalized Edema	PT 10018092 Generalised Oedema
Photosensitivity Reaction	PT 10034972 Photosensitivity Reaction
Important Potential Risks	
Pneumonitis	PTs: 10035742 Pneumonitis 10035745 Pneumonitis Chemical 10037765 Radiation Pneumonitis 10066728 Acute Interstitial Pneumonitis 10022611 Interstitial Lung Disease
Other AESIs	
Anemia	SMQ 20000029: Haematopoietic Erythropenia (broad)
Neutropenia	SMQ 20000030: Haematopoietic Leukopenia (broad)
Infusion-Related Reaction	10051792 Infusion related reaction
Thrombocytopenia	SMQ 20000031 Haematopoietic Thrombocytopenia (broad)
Haemorrhages	SMQ 20000039 Haemorrhage Terms (Excluding Laboratory Terms; narrow)
Hypoalbuminemia	PTs: 10020942 Hypoalbuminaemia 10005287 Blood Albumin Decreased 10005286 Blood Albumin Abnormal

Table 3. Adverse Events of Special Interest (Continued)

Adverse Events of Special Interest	MedDRA Search Criteria/Codes
Other AESIs	
Cutaneous Reaction	SOC 10040785 Skin and Subcutaneous Tissue Disorders The observed TEAEs under this SOC will be medically adjudicated to maintain a list of PTs that are not medically relevant and that will be excluded from the analysis. Current list of PTs to exclude: 10001760 Alopecia 10002424 Angioedema 10011985 Decubitus Ulcer 10014080 Ecchymosis 10020642 Hyperhidrosis 10051235 Madarosis 10070533 Nail Bed Disorder 10028694 Nail Disorder 10062283 Nail Ridging 10028692 Nail Discolouration 10029410 Night Sweats 10034754 Petechiae 10037087 Pruritus 10052576 Pruritus generalised 10037549 Purpura 10042682 Swelling Face
Edema	PTs: 10030095 Oedema 10030124 Oedema Peripheral 10048961 Localised Oedema 10018092 Generalised Oedema 10042674 Swelling 10016807 Fluid Retention 10016803 Fluid Overload 10024770 Local Swelling 10048959 Peripheral Swelling 10007196 Capillary Leak Syndrome 10015993 Eyelid Oedema 10016029 Face Oedema 10052139 Eye Oedema 10034545 Periorbital Oedema
Hepatotoxicity	SMQs: 20000007 Drug-Related Hepatic Disorders (severe events only; broad) 20000009 Cholestastis and Jaundice of Hepatic Origin (broad) 20000008 Liver Related Investigations, Signs and Symptoms (broad) *Exclude PT 10020942 Hypoalbuminaemia from this search since hypoalbuminemia is associated with chronic liver disease and therefore not included in AESI*

Time to onset

Time to onset is the Rx Day of the start date of the first occurrence of AESI. If a subject has not experienced an AESI, the subject will be censored on the day of the subject's last assessment (i.e., the day of the subject's last known laboratory assessment, last known vital sign assessment, last known physical exam, last known tumor assessment, or last known follow-up visit, whichever is the latest) or 70 days from the last study drug administered, whichever is earliest; if the subject has not experienced an AESI and had no post-baseline assessment (i.e., none of laboratory assessment, vital sign assessment, physical exam, or tumor assessment), the subject will be censored on the day of the subject's first study drug.

Median onset time will be estimated using a Kaplan-Meier method with corresponding 95% CI for the median for each treatment arm.

11.4 Listing of Adverse Events

The following additional summaries of adverse events will be provided:

- Listing of treatment-emergent serious adverse events
- Listing of treatment-emergent adverse events that lead to discontinuation of study drug
- Listing of treatment-emergent fatal adverse events
- Listing of Grade 3 or higher adverse events
- Listings of SAEs leading to death
- Listings of SAEs leading to treatment discontinuation.

11.5 Analysis of Laboratory, Vital Signs, and ECG Data

Clinical laboratory variables are listed in Table 4.

For laboratory (hematology, chemistry, urinalysis, and coagulation), vital signs, and ECG parameters, changes from baseline are analyzed for each post-baseline visit (as specified

in Section 6.0 Analysis of Conventions) and the final visit. If more than one measurement exists for a subject on a particular day, an arithmetic average will be calculated and considered to be the subject's measurement of that day.

Descriptive statistics will be presented for baseline, each post-baseline, and the final visit (the last valid record) by each treatment arm. Mean change from baseline to each post baseline visit, and to the final visit within each treatment arm will also be presented. The highest and lowest values of each parameter will be identified for each subject, and descriptive statistics for mean change from baseline within each treatment arm will be presented.

Laboratory grade determinations will be categorized according to NCI CTCAE version 4.0 grades, and shifts from baseline grades to maximum and final post-baseline grades will be assessed. The baseline and final grades will be defined respectively as the grade of the last measurement collected prior to or on the first dose of study drug, and as the grade of the last post-baseline measurement collected no more than 70 days after the last dose of study drug. Frequency and percentage of subjects will be presented by treatment arm: 1) baseline grade vs. maximum post-baseline grades, and 2) baseline grade vs. final post-baseline grade. Frequency and percentage of subjects experiencing a shift from baseline grades of 0 to 2 to maximum post-baseline grades of 3 to 4, and from baseline grades of 0 to 2 to final post-baseline grades of 3 to 4 will be presented by treatment arm.

Detailed listings of data for subjects experiencing NCI CTCAE grade 3 to 4 biochemistry and hematology values will be provided. All measurements collected, regardless of the number of days after the last dose of study drug, will be included in these listings. A listing will be provided for out-of-normal range as well as clinically significant abnormal lab values.

Descriptive summary statistics for categorical ECG status data will be presented for baseline and post-baseline visits. Shift tables will also be provided.

Table 4. Clinical Laboratory Tests

Hematology	Clinical Chemistry	Urinalysis (Dipstick)
Hematocrit Hemoglobin Red blood cell (RBC) count White blood cell (WBC) count Neutrophils Bands (if indicated) Lymphocytes Monocytes Basophils (if indicated) Eosinophils (if indicated) Platelet count (estimate not acceptable) Mean corpuscular volume Mean corpuscular hemoglobin concentration RBC distribution width	Blood urea nitrogen (BUN) Serum creatinine Total bilirubin Serum glutamic-pyruvic transaminase (SGPT/ALT) Serum glutamic-oxaloacetic transaminase (SGOT/AST) Alkaline phosphatase Sodium Potassium Calcium Inorganic phosphorus Uric acid Total protein Glucose Albumin Magnesium Chloride Amylase Lipase Lactate dehydrogenase (LDH)	Specific gravity Ketones pH Protein Blood Glucose Urobilinogen Bilirubin Leukocyte esterase
Coagulation		**Serum Pregnancy Test**
Activated Partial Thromboplastin Time (aPTT) Prothrombin time (PT) International Normalized Ratio (INR)		Beta-Human Chorionic Gonadotropin (β-hCG) (if applicable) Follicle-stimulating hormone (FSH) (if applicable)

11.6 Analysis of ECOG Performance Status

ECOG performance status will be summarized for each visit by treatment arm. Shifts from baseline to the best and worst post-baseline score will be tabulated.

11.7 Analysis of Vital Signs Using Criteria for Potentially Clinically Significant Vital Sign Values

Vital signs values will be assessed for potential clinical significance through the application of criteria developed by the sponsor as detailed in the table below.

Systolic Blood Pressure	> 150 mmHg and > 20 mmHg higher than baseline < 70 mmHg and a decrease of ≥ 30 mmHg from baseline
Diastolic Blood Pressure	> 100 mmHg and higher than baseline < 50 mmHg and a decrease of ≥ 20 mmHg from baseline
Pulse Rate	> 120 bpm and an increase of ≥ 30 bpm from baseline < 50 bpm and a decrease of ≥ 30 bpm from baseline
Temperature	≥ 38.9°C ≤ 35.6°C

The number and percentage of subjects with post baseline values meeting Criteria for Potentially Clinically Significant Vital Signs values will be summarized. Except for the temperature, subjects who have a baseline measurement and at least one post-baseline measurement will be included in the summary. A separate listing of all the subjects and values that meeting the criteria will be provided. No comparisons of the rates of subjects met the above criteria between the treatment arms will be performed.

12.0 References

1. Bretz F, Maurer W, Brannath W, et al. A graphical approach to sequentially rejective multiple test procedures. Stat Med. 2009;28(4):586-604.
2. Brookmeyer R, Crowley JA. Confidence Interval for the Median Survival Time. Biometrics. 1982;38:29-41.
3. Cox DR. Regression Models and Life Tables. Journal of the Royal Statistical Society. Series B. 1972;20:187-220. With discussion.
4. Aaronson N, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQC30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst. 1993;85(5):365-76.
5. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol. 1991;59(1):12-9.