STATISTICAL ANALYSIS PLAN

Study Protocol Number:	BGB-A317-A1217-302 (AdvanTIG-302)
NCT:	NCT04746924
Study Protocol Title:	A Phase 3, Randomized, Double-Blind Study of Ociperlimab, an Anti-TIGIT Antibody, in Combination With Tislelizumab Compared to Pembrolizumab in Patients With Previously Untreated, PD-L1-Selected, and Locally Advanced, Unresectable, or Metastatic Non-Small Cell Lung Cancer
Date:	08 Jan 2025
Version:	1.0

SIGNATURE PAGE

Author: [redacted] **Associate Director, Statistics and Data Science**	[redacted]

Approval

[redacted] **Vice President, Head of Global Statistics**	[redacted]
[redacted] **Vice President, Clinical Development, GI Franchise**	[redacted]
[redacted] **Senior Director, Clinical Development**	[redacted]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Term
ADA	Antidrug antibody
ADI	Actual dose intensity
AE	Adverse event
ALK	Anaplastic lymphoma kinase
ATC	Anatomical Therapeutic Chemical
AUC	Area under the curve
BGB-A317	Tislelizumab
BGB-A1217	Ociperlimab
BMI	Body mass index
BOR	Best overall response
CBR	Clinical benefit rate
CI	Confidence interval
CMH	Cochran-Mantel-Haenszel
C_{max}	Maximum observed plasma concentration
C_{min}	Minimum observed plasma concentration
CR	Complete response
CT	Computed tomography
DCR	Disease control rate
DOR	Duration of response
ECG	Electrocardiogram
ECOG PS	Eastern Cooperative Oncology Group Performance Status
eCRF	Electronic case report form
EGFR	Epidermal growth factor receptor

EORTC QLQ-C30	European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30
EORTC QLQ-LC13	European Organization for Research and Treatment of Cancer Quality of Life Questionnaire – Lung Cancer 13
EOT	End-of-Treatment
EQ-5D-5L	5-Level EuroQol 5-Dimension
GEP	Gene expression profiling
GHS	Global health status
HR	Hazard ratio
HRQoL	Health-related quality of life
IA	Interim analysis
IC	Immune cells
ICH	International Conference on Harmonization
IDMC	Independent Data Monitor Committee
imAE	Immune-mediated adverse event
IRT	Interactive Response Technology
ITT	Intent-to-Treat
KM	Kaplan-Meier
LKADT	Last Known Alive Date
MedDRA	Medical Dictionary for Regulatory Activities
MRI	Magnetic resonance imaging
MSI	Microsatellite instability
NCI-CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
NSCLC	Non-small cell lung cancer
ORR	Objective response rate

OS	Overall survival
PD	Progressive disease
PD-L1	programmed cell death ligand-1
PFS	Progression-free survival
PFS2	Progression-free survival after next line of treatment
PK	Pharmacokinetic(s)
PGI-S	Patient global impression of severity
PR	Partial response
PRO	Patient-reported outcomes
PRTSE	Patient reported treatment-related side-effect burden
PT	Preferred term
QTcF	Fridericia's correction formula
RDI	Relative dose intensity
RECIST	Response evaluation criteria in solid tumors
RMST	Restricted mean survival time
SAE	Serious adverse event
SAP	Statistical analysis plan
SD	Stable disease
SDRB	Sponsor Data Review Board
SOC	System organ class
SOP	Standard operating procedure
TA	Tumor assessment
TC	Tumor cells
TCM	Tradition Chinese Medicine

TEAE	Treatment-emergent adverse event
TIGIT	T-cell immunoglobulin and ITIM domain
TMB	Tumor mutation burden
TTD	Time to deterioration
TTR	Time to response
WHO DD	World Health Organization Drug Dictionary

1. INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to describe the procedures and the statistical methods that will be used to analyze and report results for study BGB-A317-A1217-302 (AdvanTIG-302): A Phase 3, Randomized, Double-Blind Study of Ociperlimab, an Anti-TIGIT Antibody, in Combination With Tislelizumab Compared to Pembrolizumab in Patients With Previously Untreated, PD-L1-Selected, and Locally Advanced, Unresectable, or Metastatic Non-Small Cell Lung Cancer. This SAP is based on AdvanTIG-302 Protocol Amendment 5.0, dated on 22 Dec, 2023. The focus of this SAP is for the planned interim analyses and the final analysis specified in the study protocol. The analysis details for Pharmacokinetic (PK), Pharmacodynamics, Pharmacogenomics and Biomarkers are not described within this SAP. Separate analysis plans might be completed for these analyses and will be attached in addition to this SAP to the clinical study report.

High PD-L1 expression is determined by PD-L1-stained ≥ 50% of tumor cells [TC ≥ 50%] using the Ventana PD-L1 (SP263) assay.

2. STUDY OVERVIEW

2.1. Study Design

This is a randomized, double-blind, multicenter, Phase 3 study designed to evaluate the efficacy and safety of ociperlimab in combination with tislelizumab compared with that of pembrolizumab in patients with PD-L1-selected non-small cell lung cancer (NSCLC) who have locally advanced or recurrent disease that is unresectable or not amenable to radiotherapy, with or without chemoradiotherapy, or previously untreated metastatic disease, and whose tumors do not harbor known epidermal growth factor receptor (EGFR)-sensitizing mutations, anaplastic lymphoma kinase (ALK) translocations, BRAF V600E mutations, or ROS1 mutations. The efficacy and safety of tislelizumab alone will be explored in a small cohort of the same patient population.

Approximately 660 patients will be enrolled. The final total number of patients in Arms A, B, and C is estimated to be approximately 286, 286, and 88, respectively. Blinding will be accomplished using placebo infusions of normal saline in Treatment Arms B and C so that all patients will receive 2 infusions on Day 1 of each cycle. Study treatments will be prepared by unblinded pharmacists, who will mask treatments to ensure that patients and study staff remain blinded.

At randomization, patients will be stratified by the following 2 factors:

- Regions of enrollment: Asia versus non-Asia
- Histology: Squamous versus non-squamous

Study treatments will be given as follows:

- Arm A: Tislelizumab 200 mg intravenously followed by ociperlimab 900 mg intravenously once every 3 weeks

- Arm B: Pembrolizumab 200 mg intravenously followed by placebo intravenously once every 3 weeks
- Arm C: Tislelizumab 200 mg intravenously followed by placebo intravenously once every 3 weeks

2.2. Study Assessments

Tumor response will be assessed by investigators using Response Evaluation Criteria in Solid Tumors Version 1.1 (RECIST v1.1) criteria (Eisenhauer et al 2009). Tumor imaging will be performed within 28 days before randomization. During the study, tumor imaging will be performed every 9 weeks (± 7 days) from randomization for the first 52 weeks and then every 12 weeks (± 7 days) based on RECIST v1.1. Tumor assessments should continue until disease progression is determined by the investigator. Patients who discontinue study treatment early for reasons other than disease progression (eg, toxicity) will continue to undergo tumor assessments following the original plan until the patient experiences disease progression or death, withdraws consent, is lost to follow-up, or until the study terminates, whichever occurs first.

Patient-reported outcomes (PRO) will be collected using the EORTC QLQ-C30, QLQ-LC13. and EQ-5D-5L at baseline (Day 1 of Cycle 1), every other cycle through Cycle 13, then every 4 cycles thereafter, and at the End-of-Treatment Visit; patient reported global impression of severity (PGI-S) will be collected at baseline and Cycles 5 and 7, and patient-reported treatment-related side-effect burden (PRTSE) will be collected at Cycles 5 and 7. At applicable dosing visits, PRO will be collected before any procedures or dose administration.

Patients will be evaluated for any adverse events (AEs) and serious adverse events (SAEs) occurring up to 30 days after the last dose of the study drug (all severity grades), per National Cancer Institute Common Terminology Criteria for Adverse Events [NCI-CTCAE] v5.0 or until initiation of new anticancer therapy, whichever occurs first, and for immune-mediated AEs (imAEs) occurring up to 90 days after the last dose of study drug regardless of whether or not the patient starts a new anticancer therapy. All SAEs considered related to the study drug that are brought to the attention of the investigator should be reported regardless of time since the last dose of treatment.

In this double-blind, placebo-controlled study, all patients and personnel involved in the conduct and interpretation of the study, including the investigators, BeiGene study team, and site personnel, will be blinded to the treatment assignment. Randomization data will be kept strictly confidential and will be accessible only by authorized persons per SOPs until the time of unblinding.

An Independent Data Monitor Committee (IDMC) will be established to regularly monitor the safety and efficacy of ociperlimab in combination with tislelizumab when compared with pembrolizumab plus placebo. Two interim analyses for overall survival (OS) are planned in the study, first one for futility and second one for efficacy.

3. STUDY OBJECTIVES

3.1. Primary Objective

- To compare OS between Arm A and Arm B in the Intent-to-Treat (ITT) Analysis Set.

3.2. Secondary Objective

- To compare progression-free survival (PFS) as assessed by investigators according to RECIST v1.1 between Arm A and Arm B in the ITT Analysis Set.
- To compare overall response rate (ORR) and duration of response (DOR) as assessed by investigators according to RECIST v1.1 between Arm A and Arm B in the ITT Analysis Set.
- To compare health-related quality of life (HRQoL) and time to deterioration (TTD) between Arm A and Arm B in the ITT Analysis Set.
- To further investigate the safety and tolerability of ociperlimab in combination with tislelizumab.

3.3. Exploratory Objective

- To compare disease control rate (DCR), clinical benefit rate (CBR), and time to response (TTR) as assessed by investigators according to RECIST v1.1 between Arm A and Arm B in the ITT Analysis Set.
- To evaluate OS, as well as ORR, DOR, PFS, DCR, CBR, and TTR as assessed by the investigators according to RECIST v1.1 in Arm C in the ITT Analysis Set.
- To evaluate PFS after next line of treatment (PFS2) in Arm A, Arm B and Arm C.
- To characterize the pharmacokinetic (PK) of ociperlimab and tislelizumab.
- To determine host immunogenicity to ociperlimab and tislelizumab.
- To further investigate the safety and tolerability of tislelizumab.
- To evaluate PGI-S and PRTSE in Arm A, Arm B, and Arm C in the ITT Analysis Set.
- To measure HRQoL in Arm C in the ITT Analysis Set.

4. DEFINITION OF ESTIMANDS

The primary analysis was written in estimand framework per study design as described in the protocol.

4.1. Primary Estimand – Overall Survival (OS) Benefit

The primary scientific question of interest is: Will ociperlimab in combination with tislelizumab prolong OS compared to pembrolizumab in patients with previously untreated, PD-L1 selected, and locally advanced, unresectable, or metastatic non-small cell lung cancer, regardless of whether patients received subsequent anticancer therapies?

The primary estimand is characterized by the following attributes:

1. Treatment of interest: The experimental treatment regimen is ociperlimab in combination with tislelizumab. The control treatment regimen is pembrolizumab as a single agent.
2. Population: Adult patients with PD-L1 ≥ 50% NSCLC who have locally advanced or recurrent disease that is unresectable or not amenable to radiotherapy, with or without chemoradiotherapy, or previously untreated metastatic disease, and whose tumors do not harbor EGFR sensitizing mutations, ALK translocations, BRAF V600E mutations, or ROS1 mutations.
3. Primary variable: Overall survival defined as the time from the date of randomization to the date of death due to any cause. Further details on OS are provided in Section 7.5.1.
4. Handling of intercurrent events:
    - New anticancer therapy started prior to death: Any incidence will be ignored, i.e., death or patients' data collected after the new anticancer therapy will be considered for analysis (treatment policy strategy).
    - Discontinuation of study treatment: Death or patients' data collected after the discontinuation of study treatment will be considered for analysis (treatment policy strategy).
    - Any other unforeseen intercurrent events: OS will take into account all deaths and any patients' data after any unforeseen intercurrent events.
5. Population-level summary: Hazard Ratio (HR) of OS comparing ociperlimab in combination with tislelizumab versus pembrolizumab will be estimated using Cox proportional hazard model stratified by regions of enrollment (Asia versus non-Asia) and histology (squamous versus non-squamous).

5. STUDY ENDPOINTS

5.1. Primary Endpoints

- OS (time from the date of randomization to the date of death due to any cause) in the ITT Analysis Set of Arm A and Arm B.

5.2. Secondary Endpoints

- PFS as assessed by investigators (time from the date of randomization to the date of the first objectively documented disease progression per RECIST v1.1, or death, whichever occurs first) in the ITT Analysis Set of Arm A and Arm B.
- ORR as assessed by investigators (proportion of patients with a documented, confirmed complete response (CR) or partial response (PR) per RECIST v1.1) and DOR as assessed by investigators (time from the first determination of a confirmed response per RECIST v1.1 until the first documentation of progression or death, whichever occurs first) in the ITT Analysis Set of Arm A and Arm B.
- HRQoL as assessed via PRO using the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30), its lung cancer module Quality of Life Questionnaire Lung Cancer 13 (QLQ-LC13), and the 5

Level EuroQol 5 Dimension (EQ-5D-5L) questionnaire in Arm A and Arm B. PRO endpoints include the EORTC QLQ C30's global health status/QoL (GHS), physical function and fatigue scales, and the QLQ-LC13's index score, dyspnea, coughing, hemoptysis and pain in chest, pain in arms/shoulders and peripheral neuropathy scales in the ITT Analysis Set.

- TTD, defined as time from randomization to the first occurrence of ≥ 10-point increase in scores in a worsening direction for 2 consecutive assessments or 1 assessment followed by death from any cause before the next scheduled data collection in the ITT Analysis Set.
- The incidence and severity of AEs according to NCI-CTCAE v5.0 in Arm A.

5.3. Exploratory Endpoints

- DCR (proportion of patients with confirmed CR + PR + stable disease), CBR (proportion of patients with confirmed CR + PR + durable stable disease), and TTR per RECIST v1.1 as assessed by investigators in Arm A and Arm B.
- OS, as well as ORR, DOR, PFS, DCR, CBR, and TTR per RECIST v1.1 as assessed by investigators in Arm C.
- PFS2, defined as the time from randomization to objective disease progression after next line of treatment, or death from any cause, whichever occurs first.
- Serum concentrations of ociperlimab and tislelizumab at specified timepoints.
- Immunogenic responses to ociperlimab and tislelizumab, evaluated through detection of antidrug antibodies (ADAs).
- The incidence and severity of AEs according to NCI-CTCAE v5.0 in Arm C.
- Patient-reported changes in NSCLC symptom severity from baseline via the PGI-S questionnaire and patient reported treatment-related side effect burden via the PRTSE questionnaire in Arms A, B, and C.
- HRQoL measured by changes in PROs using EORTC QLQ-C30, QLQ-LC13, and EQ-5D-5L questionnaires in Arm C.

6. SAMPLE SIZE CONSIDERATIONS

The sample size calculation is driven by the primary efficacy analyses of OS in the comparison between Arm A and Arm B in the ITT Analysis Set. Overall survival is assumed to have an exponential distribution. The 1-sided overall Type I error in the study is set at 0.025. Table 1 summarizes the statistical assumption and power in the sample size calculation. The initial number of patients planned in the study was 605 in Protocol Amendment Version 1.0 (Version 1.1 in Japan, Version 1.2.1 in Germany, and Version 1.4 in France), which included a randomization ratio of 5:5:1 (275, 275, and 55 patients in Arms A, B, and C, respectively). In order to increase the number of patients treated in the Arm C, the randomization ratio changed to 5:5:2 in Protocol Amendment Version 2.0 (Version 2.1 in Japan, Version 2.2 in Germany, and Version 2.3 in the US), and the total number of patients in the study increased by 55 patients

(from 605 to 660). The randomization ratio update is implemented after Protocol Amendment Version 2.0 is approved at each participating site (ie, Version 2.1 in Japanese sites, Version 2.2 in German sites, and Version 2.3 in US sites). It is estimated that approximately 300 patients will be randomized with a randomization ratio of 5:5:1 and approximately 360 patients will be randomized with a randomization ratio of 5:5:2. The final total number of patients in Arms A, B, and C is estimated to be approximately 286, 286, and 88, respectively. It is noted that the randomization ratio between Arm A and Arm B will remain unchanged.

Assuming an approximately 5% dropout rate (dropout hazard rate of 0.003) for OS, approximately 572 patients will be enrolled to Arms A and B in order to observe the targeted OS events at the defined time periods as shown in Table 1. The primary analyses will be performed when the target number of events is observed. Two interim analyses are planned after 65% and 80% of the total planned death events have occurred in the 2 treatment arms combined (Section 8 for more details).

Table 1: Hazard Ratio and Median OS Assumption, Number of Events, Alpha and Power in the Primary Hypothesis Tests

HR	Median in Arm A (months)	Median in Arm B (months)	Number of events	Alpha	Power
0.70	28.6	20	379	0.025	93%

Abbreviations: HR, hazard ratio; OS, overall survival

In addition, approximately 88 patients in Arm C will be enrolled to assess the antitumor activity of tislelizumab.

7. STATISTICAL METHODS

7.1. Analysis Sets

The ITT analysis set consists of all the patients who were randomized to a treatment arm. All patients will be grouped by the assigned treatment at randomization. The ITT analysis set will be used for efficacy analyses and HRQoL analyses.

The safety analysis set includes all patients who were randomized and received any dose of any study drug. Patients will be analyzed according to the study treatment they actually received, which is defined as (i) the intended treatment if it was received at least once or (ii) the first treatment received when starting therapy with study medication if the intended treatment is never received. Each patient will be classified into and analyzed consistently within one (and only one) treatment arm. The safety analysis set is used for all safety analyses.

The PK analysis set consists of all the patients who received any dose of the study drug and for whom valid study drug PK parameters can be estimated. The PK analysis set will be used for PK analyses.

The immunogenicity analysis set includes all patients who received any dose of the study drug and for whom both baseline antidrug antibodies (ADA) and at least 1 postbaseline ADA result are available.

7.2. Multiplicity Adjustment

The Type I error is strongly controlled at 0.025. OS will be tested at a 1-sided alpha of 0.025. Only when the superiority of OS is demonstrated will a full alpha of 0.025 (1-sided) be sequentially shifted to the hypothesis testing of the secondary endpoints of PFS and ORR based on the data up to the second interim analysis. The hypothesis test will be stopped at PFS if non-significant. Nominal p-values may be computed for other efficacy analyses but should be interpreted with caution.

7.3. Data Analysis General Considerations

7.3.1 Definitions and Computations

Study drugs include ociperlimab (also known as BGB-A1217), tislelizumab (also known as BGB-A317) and pembrolizumab.

Study day:

Study day will be calculated in reference to the date of the first dose of study drug for safety analysis purposes. For assessments conducted on or after the date of first dose, study day will be calculated as (assessment date –first dose date + 1). For assessments conducted before the date of first dose, study day is calculated as (assessment date – first dose date). If no dose is given, then the date of randomization will be used. There is no study day 0. In the situation where the event date is partial or missing, a partial or missing date will appear in the listings. Study day and any corresponding durations will be presented based on the imputations specified in Appendix 1.

To derive the duration of any efficacy endpoint, the reference date will be the date of randomization.

Baseline Measurements:

- For efficacy evaluation: A baseline value is defined as the last non-missing value collected prior to randomization.
- For safety evaluation: A baseline value is defined as the last non-missing value collected prior to the first dose of study drug.
- For toxicity grade of certain laboratory tests: Two baseline toxicity grades should be derived according to the directions (lower (Hypo) or higher (Hyper)). For example, a baseline hemoglobin with value between 10.0 g/dL and lower limit of normal (LLN) will have two baseline toxicity grades derived: Grade 1 for Hypo and Grade 0 for Hyper.

Study Follow-up Duration: Defined as the duration from the randomization date to the study discontinuation date (e.g. death, consent withdrawal, lost to follow-up) or to a cutoff date if the patient is still on-study.

Minimum Study Follow-up: Defined as the difference between the date of analysis cut-off and the date of last patient randomized.

All calculations and analyses will be conducted using SAS® Version 9.4 or higher.

7.3.2 Conventions

Unless otherwise specified, the following conventions will be applied to all analyses:

- 1 year = 365.25 days. Number of years is calculated as (days/365.25) rounded up to 1 significant digit.
- 1 month = 30.4375 days. Number of months is calculated as (days/30.4375) rounded up to 1 significant digit.
- Age will be calculated as the integer part of (date of informed consent – date of birth + 1)/365.25.
- P-values will be rounded to 4 decimal places. P-values that is less than 0.0001 will be presented as '< 0.0001' and p-values that is larger than 0.9999 will be presented as '> 0.9999'.
- Duration of image-based event endpoints (such as PFS) will be based on the actual date the image was obtained rather than the associated visit date.
- For laboratory results collected as < or >, a numeric value, 0.0000000001 will be subtracted or added, respectively, to the value.
- For by-visit observed data analyses, percentages will be calculated based on the number of patients with non-missing data as the denominator, unless otherwise specified.
- For continuous endpoints, summary statistics will include n, mean, standard deviation, median, Q1, Q3 and range (minimum and maximum).
- For discrete endpoints, summary statistics will include frequencies and percentages.

7.3.3 Handling of Missing Data

Missing data will not be imputed unless otherwise specified elsewhere in this SAP. Handling of missing data related to the primary estimand will be further described in Section 7.5.1. Missing dates or partially missing dates will be imputed conservatively for adverse events and prior/concomitant medications/procedures. Specific rules for the handling of missing or partially missing dates for adverse events and prior/concomitant medications/procedures are provided in Appendix 1.

By-visit endpoints will be analyzed using observed data unless otherwise specified. For observed data analyses, missing data will not be imputed, and only the observed records will be included.

7.4. Patient Characteristics

7.4.1 Patient Disposition

The number (percentage) of patients who signed informed consent, enrolled in the study, died before enrollment, and screen-failed, including those who re-screened, will be summarized. The number (percentage) of screen failure reasons will also be summarized.

The number (percentage) of patients randomized, treated, discontinued from treatment, and discontinued from the study will be summarized. The primary reason for end of study treatment

(treatment discontinuation) and end of study (study discontinuation) will be summarized by categories. Study follow up duration will be summarized descriptively.

7.4.2 Protocol Deviations

Important protocol deviation criteria will be established, and patients with important protocol deviations will be identified and documented. Important protocol deviations will be summarized for all patients in the ITT analysis set. They will also be listed by each category. Deviation categories are not mutually exclusive. Multiple deviations within the same category are counted once per patient.

7.4.3 Demographic and Other Baseline Characteristics

Demographic and other baseline characteristics will be summarized using descriptive statistics in the ITT analysis set, including the following variables:

- Age (continuously and by categories [<65 or ≥65 years])
- Sex
- Race
- Ethnicity
- Geographic region
- Weight
- BMI
- ECOG
- Tobacco use

In addition, the stratification factor of regions of enrollment: (Asia versus non-Asia) per IRT and per electronic case report form (eCRF) will be summarized for the ITT analysis set.

A listing of demographic and baseline characteristics will be provided.

7.4.4 Disease History

The number (percentage) of patients reporting disease history and characteristics data, as recorded on the eCRF, will be summarized for the ITT analysis set. Disease characteristics include:

- Disease stage at study entry
- Disease stage at initial diagnosis
- Patients with metastatic disease at study entry (yes or no)
- Time from initial diagnosis to randomization date
- Time from initial diagnosis of metastatic disease to randomization date
- Known metastatic sites at study entry
- Histologic grade

- PD-L1 expression by central testing (>=50%, < 50%, and unknown)

In addition, the stratification factor of histology (squamous versus non-squamous) per IRT and per eCRF will be summarized for the ITT analysis set.

A listing of disease history will be provided.

7.4.5 Prior Anticancer Drug Therapies and Surgeries

Prior anticancer drug therapies, prior anticancer surgeries with therapeutic intent, and prior anticancer radiotherapy will be summarized for the ITT analysis set. The therapies and surgeries with the same sequence/regimen number are counted as one prior therapy/surgery.

7.4.6 Prior and Concomitant Medications

Prior medications are defined as medications that stop before the day of the first dose of the study drug. Concomitant medications will be defined as medications that 1) started before the first dose of the study drug and were continuing at the time of the first dose of the study drug, or 2) started on or after the date of the first dose of the study drug up to 30 days after the patient's last dose (as of the 30-day Safety Follow-up).

Prior and concomitant medications will be coded using the version of the World Health Organization Drug Dictionary (WHO DD) drug codes Version 25.0 or higher. They will be further classified to the appropriate Anatomical Therapeutic Chemical (ATC) code.

The number (percentage) of patients reporting prior and concomitant medications will be summarized by ATC medication class and WHO DD preferred name in the ITT analysis set. A listing of prior and concomitant medications will be provided.

7.4.7 Medical History

Medical History will be coded using Medical Dictionary for Regulatory Activities (MedDRA) Version 26.0 or higher. The number (percentage) of patients reporting a history of any medical condition, as recorded on the eCRF, will be summarized by system organ class and preferred term for the ITT analysis set. A listing of medical history will be provided.

7.5. Efficacy Analysis

7.5.1 Primary Efficacy Endpoints

Primary estimands are defined in Section 4. Details of the statistical methods used in OS derivation and analysis are provided in this section including pre-defined sensitive analyses of the primary estimand and supplementary analyses of OS.

Variable

Overall survival is defined as the time from the randomization date to the documented death date for patients who died prior to or on the clinical cutoff date. For patients who are alive by the clinical cutoff date, OS will be censored at the last known alive date (LKADT). The LKADT will be defined as either the clinical data cutoff date for patients who are still on treatment, or last available date showing patients alive or cut-off date whichever comes first for other alive patients.

Every effort should be made to ensure complete death dates. In the rare case that the day of death date is missing, the death date will be imputed as the max (last available date showing patients alive + 1, first day of year/month of death date). A patient with an imputed death date will be considered as an event for OS analysis purposes. Death date with missing month and/or year will not be imputed for OS analysis purposes, will not be considered an OS event, and will be censored at LKADT.

Efficacy Analysis For OS:

OS in ITT Analysis Set

The null and alternative hypotheses to be evaluated for OS are:

$$H_0: OS_A \leq OS_B,$$

$$H_a: OS_A > OS_B,$$

where OS_A and OS_B represent the OS in Arm A and Arm B. The null hypothesis will be tested using a log-rank test stratified by IRT recorded region (Asia versus non-Asia) and histology (squamous versus non-squamous). If at the time of efficacy IA or final analysis the one-sided p-value is less than the corresponding stopping boundary specified in Table 4, it will be concluded that the null hypothesis is rejected and the superiority of Arm A over Arm B in OS is demonstrated at the significance level.

The HR and its 2-sided 95% CI will be estimated from a stratified Cox regression model with the same stratification factors above. In this analysis, the baseline hazard function will be allowed to vary across strata. SAS PHREG procedure with TIES=Efron option will be used to carry out this analysis in which the model statement will include the treatment group variable as the only covariate, and the STRATA statement will include the stratification variable as obtained via IRT. The distribution of OS, including median, Q1 and Q3, and event-free rates at every 3 months including 12 and 24 months, will be estimated using the Kaplan-Meier method for Arm A and Arm B. Ninety-five percent CIs for median and Q1 and Q3 of OS will be estimated using the method of Brookmeyer and Crowley (Brookmeyer and Crowley, 1982), and 95% CIs for event-free rates will be estimated using Greenwood's formula (Greenwood, 1926).

Sensitivity and supplementary analyses

Sensitivity analysis 1: "Unstratified OS analysis": To assess the impact of stratification factors, OS will be analyzed using the unstratified Cox model, and the treatment effect between Arm A and Arm B will be summarized by the hazard ratio with its 95% confidence interval.

Sensitivity analysis 2: "OS analysis with stratification factors from the eCRF": OS will be analyzed using a stratified Cox model with stratification factors obtained from the eCRF, and the treatment effect between Arm A and Arm B will be summarized by the hazard ratio with its 95% confidence interval.

Supportive analysis 3: "OS analysis to check proportional hazard assumption" Analyses to assess proportional hazard assumption, including Schoenfeld residual plot and time-dependent covariate in the Cox model, will be explored.

Supplementary analyses

Supplementary analysis 1 "OS analysis based on Max-Combo method": This analysis targets an estimand which has the same attributes as the primary estimand except the population level summary will be Weighted hazard ratio of OS between Arm A and Arm B (combo of G(0,0), G(0,1), G(1,0) and G(1,1)) from Max-combo test to account for the possible non-proportional hazards effects. (Satrajit R, Keaven A, Jiabu Y, Pralay M, 2019).

Supplementary analysis 2 "OS analysis based on Restricted mean survival time method": This analysis targets an estimand that has the same attributes as the primary estimand, except the population level summary will be the difference in the restricted mean survival time (RMST) between Arm A and Arm B. In order to account for the possible non-proportional hazard effect, the RMST (Uno H, Claggett B, Tian L, Inoue E, et al. 2014) will be computed for OS separately using the area under the curve from baseline to the minimum of the largest observed time on Arm A and Arm B and the difference with its 95% CI will be displayed.

If patients in the US and Japan are enrolled based on local PD-L1 testing, supplementary analyses of OS may be carried out using a stratified Cox regression model among all randomized patients with high PD-L1 expression (PD-L1 positive tumor cells ≥ 50%) as per the central laboratory using the VENTANA PD-L1 (SP263) CDx Assay.

7.5.2 Secondary Efficacy Endpoints

PFS by Investigator in ITT Analysis Set

PFS by investigator is defined as the time from the randomization date to disease progression as assessed by investigator per RECIST v1.1, or death, whichever occurs first. PFS will be censored at the last adequate tumor assessment if one of the following occurs by the time of analysis: Absence of event, the event occurred after a new anticancer therapy is initiated, or the event occurred after two or more missing tumor assessments. The censoring rules for the primary analysis of PFS by investigator are presented in Table 2. The algorithm to identify missing tumor assessments is presented in Appendix 2.

Table 2: Censoring rules for Progression-free Survival by Investigator Per RECIST Version 1.1

	Derivation rules	Outcome
No baseline disease assessments	Date of randomization	Censored
No post-baseline disease assessments, and no death within two scheduled disease assessments	Date of randomization	Censored
No post-baseline disease assessment, and death within two scheduled disease assessment	Date of death	Event
No PD or death by data cutoff or withdrawal from study or lost to follow up	Date of last adequate disease assessment by data cutoff or withdrawal from study or lost to follow up	Censored

New anticancer therapy started prior to PD or death	Date of last adequate disease assessment before the new anticancer therapy. If also have ≥ 2 consecutive missed disease assessments, the earlier of the last adequate disease assessment prior to missed disease assessments and before new anticancer therapy.	Censored
No new anticancer therapy; PD or death after ≤ 1 missed disease assessment	Date of PD or death, whichever occurred first	Event
No new anticancer therapy; PD or death after ≥ 2 consecutive missed disease assessments	Date of last adequate disease assessment prior to missed disease assessments	Censored

The methods used to analyze OS will be applied to the analysis for PFS. The analysis of PFS as assessed by investigators will be a log-rank test stratified by IRT recorded region (Asia versus non-Asia) and histology (squamous versus non-squamous). The HR and its 2-sided 95% CI will be estimated from a stratified Cox regression model with the same stratification factors above. The distribution of PFS, including median, Q1 and Q3, and event-free rates at every 3 months including 6 and 12 months will be estimated using the Kaplan-Meier method for Arm A and Arm B. Ninety-five percent CIs for median and Q1 and Q3 of PFS will be estimated using the method of Brookmeyer and Crowley (Brookmeyer and Crowley, 1982), and 95% CIs for event-free rates will be estimated using Greenwood's formula (Greenwood, 1926).

One sensitivity analysis is planned:

- The sensitivity analysis is the same as the primary analysis for PFS except for including any tumor assessments/death after more than one missing visit when deriving PFS. Missing more than one visit is not considered a reason for PFS censoring. This analysis is to address the impact of PFS comparison due to missing tumor assessments.

One supplementary analysis is planned:

- This analysis is the same as the primary analysis for PFS except excluding the start of new anticancer therapy as a reason for PFS censoring. Any tumor assessments after the start of new anticancer therapy, including PD or death, will be considered when deriving PFS. This analysis is to address the impact of the new anticancer therapy received prior to progression.

Objective Response Rate (ORR) by Investigators

ORR by investigator is the percentage of patients whose best overall response recorded from randomization until data cutoff, progressive disease, or start of a new anticancer treatment, is confirmed complete response (CR) or partial response (PR) assessed by investigator per RECIST v1.1. The null hypotheses of no difference in ORR by investigator between Arm A and Arm B will be tested in a Cochran-Mantel-Haenszel (CMH) test adjusting for stratification factors of IRT recorded region (Asia versus non-Asia) and histology (squamous versus non-squamous) in the ITT Analysis Set. Patients with no postbaseline response assessment (for any reason) will be considered non-responders. The 2-sided 95% CIs for the odds ratio in ORR will be calculated, as well as Clopper- Pearson 95% CIs of ORR, for Arm A and Arm B.

Duration of Response (DOR) by Investigators

For patients with confirmed CR or PR assessed by investigator, DOR is defined as progression event free or alive time counted from the first confirmed objective response date to the first documented radiological PD date or death date, whichever occurred first. All the censoring rules for PFS by investigator should be applied to DOR. DOR assessed by investigators will be analyzed in the responders only. The median DOR and the cumulative probability of DOR estimated every 3 months will be calculated using Kaplan-Meier estimates in Arm A and Arm B and presented with 2-sided 95% CIs using the method of Brookmeyer and Crowley (Brookmeyer and Crowley, 1982). No formal testing will be performed to compare DOR between two treatment groups as it would be based on a non-randomized subgroup.

Health-Related Quality of Life

The EORTC QLQ-C30 (QLQ-C-30) consists of thirty questions that are specific to cancer and cancer treatment (Aaronson NK, et al., 1993; Fayers PM, et al., 2001). It includes a global health status (GHS/QoL) scale that consists of 2 items and five functional scales measuring Physical (5 items), Role (2 items), Cognitive (2 items), Emotional (4 items), and Social (2 items), with higher scores, indicating better HRQoL. Also, three symptom scales measuring Fatigue (3 items), Pain (2 items), and Nausea and Vomiting (2 items) and six single items measuring Dyspnoea, Insomnia, Appetite Loss, Constipation, Diarrhea, and Financial Difficulties. QLQ-C30 scores are based on 4-point Likert scales from 1 = "Not At All" to 4 = "Very Much" with lower score indicating better HRQoL; except for the GHS/Qol that is scored on a 7-point scale ranging from 1 = "Very Poor" to 7 = "Excellent".

QLQ-LC13 is a lung cancer module of EORTC-QLQ-C30 and consists of 13 items that comprises 10 lung cancer-specific scales with Dyspnoea (3 items) and 1-tem scales including, Haemoptysis, Sore mouth, Dysphagia, Peripheral neuropathy, Alopecia, Pain in chest, Pain in arm or shoulder, Pain in other parts. Scores are based on 4-point Likert scales from 1 = "Not At All" to 4 = "Very Much" with lower score indicating better HRQoL/symptomology.

EQ-5D-5L – The instrument comprises a descriptive module and a Visual Analogue scale (VAS). The descriptive module comprises of 5 dimensions: mobility, self-care, usual activities, pain/ discomfort, and anxiety/depression. Each dimension has 5 levels: 1= no problems, 2 = slight problems, 3 = moderate problems, 4 = severe problems, and 5 = extreme problems. Higher scores indicate lower quality of life. The EQ VAS measures respondent's self-rated health status on a 0 to 100 scale, with 100 = 'the best health you can imagine' and 0 = 'the worst health you can imagine'. Higher scores on VAS indicate higher health status. Descriptive analysis for VAS score will be performed and reported.

PGI-S includes one item regarding the severity of the LC symptoms with a 5-Likert scale scoring with 1=not at all and 5=extremely. Higer scores indicates worse outcomes.

PRTSE includes one item regarding the treatment burden with a 5-Likert scale scoring with 1=not at all and 5=extremely. Higer scores will show higher levels of burden.

EORTC Scoring Derivation

The principle for scoring applies to all scales/scores. The derived scales are obtained from the raw scores as defined in the EORTC manual. If at least half of the items for a scale are answered, then

the remaining completed items are used to calculate the score for that scale; however, if more than half are missing, the scale score is set to missing.

Raw Score (RS)

Raw scores are calculated as the average of the items that contribute to the scale:

$$RS=(I1+I2+\ldots+ In)/n$$

Derived Scale (S)

A linear transformation to standardize the raw scores is utilized, so that the scores are ranged from 0 to 100. The derived scales have a more intuitive interpretation: higher scores in functional scales and the global health status/QoL(GHS/Qol) indicate improvements while higher scores in symptom scales and items indicate deteriorations. The derivation formulas are computed as follows:

Functional scales:

$$S=[1- (RS-1)/range]*100$$

Symptom scales and global health status:

$$S=[(RS-1)/range]*100$$

Scales/Items: QLQ-C30

GHS and Functional Scales: Higher scores = Better HRQoL

Symptom Scales: Lower Scores = Better HRQoL

	Scale	Number of items	Item range	Item Numbers
Global health status/ QoL Global health status/QOL	QL2	2	6	29,30
Functional Scales				
Physical functioning	PF2	5	3	1, 2, 3, 4, 5
Role functioning	RF2	2	3	6, 7
Emotional functioning	EF	4	3	21, 22, 23, 24
Cognitive functioning	CF	2	3	20, 25
Social functioning	SF	2	3	26, 27
Symptom Scales/ items				
Fatigue	FA	3	3	10, 12, 18
Nausea and vomiting	NV	2	3	14, 15
Pain	PA	2	3	9, 19

Dyspnoea	DY	1	3	8
Insomnia	SL	1	3	11
Appetite loss	AP	1	3	13
Constipation	CO	1	3	16
Diarrhoea	DI	1	3	17
Financial Difficulties	FI	1	3	28

Scales/Items: QLQ-LC13

Lower Scores = Better HRQoL

	Scale	Number of items	Item range	Item Numbers
Symptom scales/items				
Dyspnea	LCDY	3	3	33,34,35
Coughing	LCCO	1	3	31
Hemoptysis	LCHA	1	3	32
Sore mouth	LCSM	1	3	36
Dysphagia	LCDS	1	3	37
Peripheral neuropathy	LCPN	1	3	38
Alopecia	LCHR	1	3	39
Pain in chest	LCPC	1	3	40
Pain in arm or shoulder	LCPA	1	3	41
Pain in other parts	LCPO	1	3	42
Did you take any medicine for pain? If Yes, how much did it help?		1	1 3	43

Descriptive analysis will be performed for all the scales for the three arms.

A mixed effect model analysis will be performed to assess clinically meaningful changes from baseline (95%CI) in the PRO endpoints (GHS and physical function scales, and symptoms of fatigue measured by QLQ-C30, and dyspnea, coughing, hemoptysis, pain in chest, pain in arms/shoulders and peripheral neuropathy, and the index score of QLQ-LC13) at key assessment timepoints, ie, Cycles 5 and 7 for all Treatment Arms A, B and C. The model will include

baseline score, stratification factors, treatment arm, visit, and treatment arm by visit interaction as fixed effects and visit as a repeated measure with an unstructured covariance structure. Clinically meaningful change for the EORTC PRO instruments has been defined as 5-point (5%) change in the scores (Osoba 1998). LS mean change (95%CI) with descriptive p values will be calculated to estimate the differences between the Arm A and Arm B.

As exploratory analysis to measure contribution of compound, LS mean change (95%CI) with descriptive p values will be calculated to estimate the differences between the Arm A and Arm C.

As exploratory analyses, for PGI-S and PRTSE, changes from baseline to cycles 5 and 7 will be summarized descriptively.

Time to Deterioration (TTD)

TTD will be analyzed using PRO endpoints. The deterioration threshold is defined as the time from randomization to the first occurrence of a ≥ 10-point increase in scores in a worsening direction. A deterioration was not counted as an event if an improvement subsequently occurred. A nonparametric Kaplan-Meier method will be used to estimate the deterioration curve in each group. A Cox regression model will be estimated for Arm A and B that included treatment as a covariate, regions of enrollment: (Asia versus non-Asia) and histology (squamous versus non-squamous) as stratification factors. The stratified log-rank test and stratified hazard ratio will be provided to show the magnitude of treatment effects, and P-values will be calculated for descriptive purposes.

7.5.3 Subgroup Analyses

To assess whether the treatment effect is consistent across various subgroups in Arm A and Arm B, the median OS in each of the subgroups along with unstratified OS hazard ratio and their 95% CIs will be estimated and plotted within each category of the following variables:

- Tegion (Asia versus non-Asia)
- Histology (squamous versus non-squamous)
- Race
- Age (< 65 versus >= 65 years)
- Gender (female versus male)
- ECOG PS (0 versus 1)
- Presence of brain metastases at study entry (yes or no)
- Presence of liver metastases at study entry (yes or no)
- Stage at study entry (locally advanced [<=III] versus metastatic [IV])
- Previous systemic anticancer therapy (yes or no)
- Smoking status (former/current versus never)
- PD-L1 expression by central testing (>=50%, < 50%, and unknown)

Country-specific subgroups may also be summarized per local regulatory requirements.

7.5.4 Exploratory Efficacy Endpoints

Best Overall Response (BOR) by Investigator

BOR by investigator is defined as the best response assessed by the investigators per RECIST v1.1 recorded from randomization until data cutoff, progressive disease, or start of a new anticancer treatment. The proportion and its corresponding Clopper-Pearson 95% CI for each of the response categories (CR, PR, stable disease (SD), and progressive disease (PD)) will be presented for Arm A and Arm B.

Disease Control Rate (DCR), Clinical Benefit Rate (CBR), Time to Response (TTR) by Investigator

DCR is defined as the proportion of patients whose BOR is CR, PR, or SD including non-CR/non-PD. CBR is defined as the proportion of patients who have CR, PR, or SD including non-CR/non-PD of ≥ 24 weeks in duration. DCR and CBR by investigators will be analyzed similarly to ORR by investigator in Arm A and Arm B.

TTR will be summarized using descriptive statistics, such as mean, median, and standard deviation. Only patients who have achieved an objective response will be included in the analysis of TTR in Arm A and Arm B.

Waterfall plots will be provided for the maximum tumor shrinkage based on the change in the target lesion sum of diameters from baseline. In addition, patients that had a tumor reduction in target lesions but overall response of PD due to new-lesions or progression of non-target lesions will be flagged in the plot. The maximum tumor shrinkage based on the change in the target lesion sum of diameters used in the plots will be listed. These analyses will be performed based on RECIST v1.1 by investigator.

OS, PFS, TTR, DOR, ORR, DCR, and CBR in Arm C

ORR, DCR, and CBR as assessed by investigator with a Clopper-Pearson 95% CI will be summarized in patients receiving tislelizumab (Arm C). The odds ratio in ORR between Arm A and Arm C with 2-sided 95% CIs will be calculated.

The distribution of OS, DOR, and PFS as assessed by investigator in patients receiving tislelizumab will be analyzed based on the Kaplan-Meier method. The median and 95% CI using the method of Brookmeyer and Crowley will be calculated. The HRs in OS and PFS between Arm A and Arm C, between Arm B and Arm C, and their 2-sided 95% CIs will be estimated from Cox regression model. Odds ratio in ORR between Arm A and Arm C; between Arm B and Arm C and their 2-sided 95% CIs will be calculated, as well as Clopper- Pearson 95% CIs of ORR.

TTR will be summarized descriptively.

Progression-free survival after next line of treatment (PFS2)

PFS2 is defined as the time from the randomization date to the first documented disease progression on next-line therapy or death from any cause, whichever occurs first. The first documented progression on next-line treatment will be recorded by investigator (i.e. captured on the post treatment discontinuation anti-cancer systemic therapy eCRF page).

- Next-line therapy is defined as the first new (systemic) anti-neoplastic therapy initiated after discontinuation of study treatment regardless of end of treatment (EOT) reason. Drugs given as part of the same regimen should be grouped as one line (i.e. part of the next-line therapy). In addition, continuation of the study treatment after the initial radiologic disease progression will not be considered as next-line therapy.
- PFS2 will be censored if no PFS2 event (progression or death) is observed during next-line therapy before the analysis cut-off date; the censoring date will be the date of last known to be alive.
- In case a second new anti-cancer therapy is introduced without progression on the first next anti-cancer therapy, then PFS2 will be censored at the end date of the first new anti-cancer therapy (i.e. next line therapy).
- PFS2 will be censored at the date of last known to be alive. If a patient is still ongoing on study treatment irrespective of the disease progression status or second progression while being on study treatment, or patient has discontinued study treatment but has not started next-line therapy and is still alive.
- Any death prior to initiation of next-line therapy will be considered as an event for PFS2 purposes.

Kaplan-Meier (KM) method as described in the PFS analyses will be used in the analysis of PFS2. The median PFS2 and the cumulative probability of PFS2 estimated at every 3 months will be calculated using Kaplan-Meier estimates for Arm A and Arm B and presented with 2-sided 95% CIs computed by Brookmeyer and Crowley method using the log-log transformation. The hazard ratio (HR) of PFS2 from stratified Cox model will be estimated and presented with a 2-sided 95% CI.

7.5.5 Post and During-treatment Anti-Cancer Therapy

Post treatment anti-cancer therapy is defined as the anti-cancer therapy started after the last dose of study drug(s). A summary of number and percentage of patients who received subsequent systematic anticancer therapy/immune checkpoint inhibitors (single treatment), and combination therapy of immune checkpoint inhibitors and tyrosine kinase inhibitors will be provided by arm based on ITT analysis set and PD-L1 positive analysis set.

Separate flags of start date of new anti-cancer therapy for efficacy and safety analyses are derived individually.

- As for efficacy analysis, the start date of new anti-cancer therapy will be the earliest date of prohibited anti-cancer therapy taken during treatment, the date of the post-treatment systemic anti-cancer therapy, and the date of other anti-cancer therapy such as post-treatment surgery and radiotherapy as deemed appropriate.
- The start date of new anti-cancer therapy in defining TEAE for safety is always the first date of new systemic anti-cancer therapy taken after the last study treatment.

Tumor response per RECIST or event-driven endpoints have not been commonly used for the efficacy evaluation of Tradition Chinese Medicine (TCM). ORR, PFS or OS benefit of Chinese

herbal medicines or Chinese patent medicines has not yet been established. Therefore, they will not be taken into account as new anti-cancer therapy in the efficacy and safety analyses.

Patient data listings of post-treatment anti-cancer therapy, procedure, radiotherapy, or surgery will be provided.

7.6. Safety Analyses

All safety analyses will be performed by treatment arms based on the safety analysis set. Safety and tolerability will be assessed, where applicable, by incidence, severity, and change from baseline values for all relevant parameters including all adverse events graded by NCI-CTCAE v5.0, laboratory values (e.g., hematology, clinical chemistry), vital signs, ECGs and physical examination. Descriptive statistics (e.g., n, mean, standard deviation, median, Q1, Q3, minimum, maximum for continuous variables; n [%] for categorical variables) will be used to analyze all safety data in the safety analysis set.

7.6.1 Extent of Exposure

The following measures of the extent of exposure will be summarized with descriptive statistics for each study drug. One cycle is defined as 21 days of treatment. Specifically:

- Duration of exposure (days): Defined as the duration from the date of the first dose of the study drug to the date of the last dose of the study drug (last date of exposure – date of first dose + 1)
    - If patients discontinued treatment (with non-missing EOT date), use min (cutoff date, study discontinuation date, death date, last dose date + 20) as the “last date of exposure.”
    - Otherwise for treatment ongoing patient, using cutoff date as the “last date of exposure”
- Number of treatment doses received (number and percentage of patients): Defined as the total number of treatment doses per administration at all visits prior to the cutoff date.
- Cumulative total dose received per patient (mg): Defined as the cumulative dose of the study drug during the treatment period of the study.
- Actual dose intensity (ADI) (mg/cycle): Defined as the cumulative total dose received by a patient divided by the duration of exposure in cycles. 21*total cumulative dose (mg) / (last dose date prior to cutoff date+ 21 - first dose date).
- Relative dose intensity (RDI) (%): Defined as the ratio of the actual dose intensity and the planned dose intensity. Planned dose intensity is defined as the planned dose on study day 1 by a patient divided by the duration of exposure in cycles.
    - For tislelizumab and pembrolizumab: RDI (%) = ADI/200*100%
    - For ociperlimab: RDI (%) = ADI/900*100%
- Number (%) of patients with dose delay
- Number (%) of patients with dose interruption

- Number (%) of patients with dose discontinuation
- Reasons for dose delay
- Reasons for dose discontinuation
- Reasons for dose interruptions

Patient data listings will be provided for all dosing records and for calculated summary statistics.

7.6.2 Adverse Events (AEs)

AEs will be graded by the investigators using NCI-CTCAE v5.0. The AE verbatim descriptions (investigator terms from the eCRF) will be classified into standardized medical terminology using MedDRA. Adverse events will be coded to the MedDRA (Version 26.0 or higher) lowest level term closest to the verbatim term, along with the linked MedDRA preferred term (PT) and primary system organ class (SOC).

A treatment-emergent adverse event (TEAE) is defined as an AE that had onset or increase in severity level date on or after the date of the first dose of study drug through 30 days after the last dose of study drug or the initiation of new anti-cancer therapy, whichever is earlier. Only those AEs that were treatment emergent will be included in summary tables of TEAE.

Immune-mediated AEs (imAEs) will be identified from all AEs that had an onset date or a worsening in severity from baseline (pretreatment) on or after the first dose of study drug and up to 90 days from the last dose of study drug, regardless of whether the patient starts a new anti-cancer therapy. If an imAE occurs outside of the above mentioned TEAE window, it will not be classified as a TEAE. All imAEs will be reported separately.

All AEs, treatment-emergent or otherwise, will be presented in patient data listings.

7.6.2.1. Treatment Emergent Adverse Event (TEAE)

Summaries of the following TEAEs will be provided in the safety analysis set:

- All TEAEs
    - Overview of TEAE
    - All TEAEs by PT
    - All TEAEs by SOC and PT (any grade and ≥ grade 3)
    - All TEAEs by SOC, PT and worst grade
    - Treatment-related TEAEs by PT
    - Ociperlimab/Placebo related TEAEs by PT
    - Treatment-related TEAEs by SOC and PT (any grades and ≥ grade 3)
    - Ociperlimab/Placebo related TEAEs by SOC and PT (any grades and ≥ grade 3)
    - Tislelizumab/Pembrolizumab related TEAEs by SOC and PT (any grades and grade ≥ grade 3)
    - Most frequently reported (incidence ≥ 3% in any treatment arm) TEAEs by PT

  - Most frequently reported (incidence ≥ 3% in any treatment arm) ociperlimab/placebo related TEAEs by PT
  - Most frequently reported (incidence ≥ 3% in any treatment arm) treatment-related TEAEs by PT
- Serious TEAEs
  - Serious TEAEs by PT
  - Serious TEAEs by SOC and PT
  - Treatment-related serious TEAEs by SOC and PT
  - Ociperlimab/Placebo related serious TEAEs by SOC and PT
  - Tislelizumab/Pembrolizumab related serious TEAEs by SOC and PT
  - Most frequently reported (incidence ≥ 3% in any treatment arm) serious TEAE by PT
- TEAEs leading to death
  - TEAEs leading to death by PT
  - TEAEs leading to death by SOC and PT
  - Treatment-related TEAEs leading to death by SOC and PT
  - Ociperlimab/Placebo related TEAEs leading to death by SOC and PT
  - Most frequently reported (incidence ≥ 3 % in any treatment arm) TEAEs leading to death by PT
- TEAEs leading to treatment discontinuation
  - TEAEs leading to treatment discontinuation by PT
  - TEAEs leading to treatment discontinuation by SOC and PT
  - Treatment-related TEAEs leading to treatment discontinuation by SOC and PT
  - TEAEs leading to Ociperlimab/Placebo discontinuation by SOC and PT
  - Treatment-related TEAEs leading to Ociperlimab/Placebo discontinuation by SOC and PT
  - TEAEs leading to Tislelizumab/Pembrolizumab discontinuation by SOC and PT
  - Treatment-related TEAEs leading to Tislelizumab/Pembrolizumab discontinuation by SOC and PT
  - Most frequently reported (incidence ≥ 3 % in any treatment arm) TEAE leading to treatment discontinuation by PT
- TEAEs leading to treatment modification including dose delay, dose interruption, and infusion rate decrease
  - TEAEs leading to treatment modification by PT
  - TEAEs leading to treatment modification by SOC and PT

- Treatment-related TEAEs leading to treatment modification by SOC and PT
- TEAEs leading to treatment modification of Ociperlimab/Placebo by SOC and PT
- Treatment-related TEAEs leading to treatment modification of Ociperlimab/Placebo by SOC and PT
- TEAEs leading to treatment modification of Tislelizumab/Pembrolizumab by SOC and PT
- Treatment-related TEAEs leading to treatment modification of Tislelizumab/Pembrolizumab by SOC and PT
- Most frequently reported (incidence ≥ 3 % in any treatment arm) TEAEs leading to treatment modification by PT

Most frequent TEAEs by PT tables will be presented in in-text tables only.

7.6.2.2. Immune-mediated Adverse Event (imAE)

Immune-mediated adverse events (imAEs) are of special interest and will be summarized by category within a pre-defined list. The identification of imAEs is described in the immune-mediated adverse event charter. All imAEs up to 90 days after the last dose of study drug will be summarized.

Summaries of the following incidences of imAEs will be provided in all treatment arms:

- Overview of imAEs
- imAEs by category and PT (any grade and ≥ grade 3)
- imAEs by category, PT and worst grade
- Serious imAEs by category and PT
- imAEs leading to death by category and PT
- imAEs leading to treatment discontinuation by category and PT
- imAEs leading to treatment discontinuation of Ociperlimab/Placebo by category and PT
- imAEs leading to treatment discontinuation of Tislelizumab/Pembrolizumab by category and PT
- imAEs leading to treatment modification by category and PT
- imAEs leading to treatment modification of Ociperlimab/Placebo by category and PT
- imAEs leading to treatment modification of Tislelizumab/Pembrolizumab by category and PT
- imAEs treated with systematic corticosteroid by category
- imAEs treated with other immunosuppressant by category
- imAEs treated with hormone therapies by category

7.6.2.3. Infusion-related Reactions

For infusion related reactions (IRR), a summary of IRRs and IRR incidence by SOC and PT (all grades and ≥ grade 3) will be provided, sorted by descending order of incidence within each SOC and PT based on Arm A.

7.6.3 Deaths

All deaths and causes of death will be summarized by treatment group, including those that occurred during the study treatment period and those reported during the survival follow-up period after treatment completion/discontinuation.

7.6.4 Laboratory Values

Laboratory safety tests will be evaluated for selected parameters described in Table 3.

Laboratory parameters that are graded in NCI CTCAE Version v5.0 will be summarized by shifts from baseline CTCAE grades to maximum post-baseline grades. In the summary of laboratory parameters by CTCAE grade, parameters with CTCAE grading in both high and low directions will be summarized separately.

Box-whisker plots will be generated for parameters of interest.

Hy's Law for liver injury will also be summarized.

Table 3: Clinical Laboratory Assessments

Serum chemistry	**Hematology**	**Coagulation**	**Urinalysis (at Screening and as clinically indicated)**
Alkaline phosphatase	Hemoglobin	Prothrombin time	pH
Alanine aminotransferase	Hematocrit	Partial thromboplastin time or activated partial thromboplastin time	Specific gravity
Aspartate aminotransferase	White blood cell count	International normalized ratio	Glucose
Albumin	Neutrophil count		Protein
Total bilirubin	Lymphocyte count		Ketones
Direct bilirubin	Platelet count		Blood
Blood urea nitrogen or urea			24-hour protein
Potassium			
Sodium			
Calcium			
Creatinine			
Glucose			
Lactate dehydrogenase			

Serum chemistry	Hematology	Coagulation	Urinalysis (at Screening and as clinically indicated)
Total protein			
Magnesium			
Phosphorus			
Chloride			
Creatine kinase/CK-MB			

Abbreviations: CK-MB, creatine kinase cardiac muscle isoenzyme.

7.6.5 Vital Signs, Height and Weight

Descriptive statistics for vital sign parameters (body temperature, pulse rate, and systolic and diastolic blood pressure), weight and changes from baseline will be presented by visit. Height will be recorded at screening only. The change from post-dose (end of infusion) to pre-dose also need to be summarized for all vital sign parameters. Vital signs will be listed by patient and visit. Box-whisker plots will be generated for actual value and change from baseline for systolic and diastolic blood pressure.

7.6.6 Electrocardiograms (ECG)

12-lead ECG recordings are required at Screening, Safety Follow-up, and as clinically indicated. Patient listing of ECG will be provided for all ECG recordings. The actual value and the change from baseline for QTcF intervals will be summarized by visit and treatment group using descriptive statistics. Abnormal post-baseline QTcF results will be summarized with the following categories: increase of >30 msec, increase of > 60 msec, value of > 450 msec, value of > 480 msec, value of > 500 msec for each visit by treatment group.

7.6.7 ECOG

ECOG performance status will be summarized by treatment arm and by visit. A shift table from baseline to worst post-baseline in ECOG performance score will be summarized.

7.6.8 Immunogenicity Analyses

Samples to assess anti-ociperlimab and anti-tislelizumab antibodies will be collected only in patients who receive study drugs and at sites that are able to adequately perform sampling, handling, and processing as outlined in the laboratory manual.

ADA attributes:

- **Treatment-boosted ADA** is defined as ADA positive at baseline that was boosted to a 4-fold or higher level following drug administration.
- **Treatment-induced ADA** is defined as ADA not detected in the baseline sample, but ADA detected after treatment administration.
- **Transient ADA response** is defined as treatment-induced response that is not considered persistent

- **Persistent ADA response** is defined as treatment-induced ADA detected at 2 or more time points during treatment (including follow-up period if any), where the first and last ADA positive samples (irrespective of any negative samples in between) are separated by 16 weeks or longer or treatment induced ADA detected in the last sampling time point.
- **Neutralizing ADA** is defined as ADA that inhibits or reduces the pharmacological activity.

ADA response endpoints:

- **ADA incidence** is defined as the sum of treatment-induced and treatment-boosted ADA-positive patients as a proportion of the ADA evaluable population.
- **ADA prevalence** is defined as the proportion of all patients that are ADA positive, including pre-existing ADA, at any time point.

The immunogenicity results will be summarized using descriptive statistics by the number and percentage of patients who develop detectable ADAs. The incidence of positive ADAs and neutralizing ADAs will be reported for evaluable patients.

7.6.9 Other Safety Measurements

Other safety measurements, including pulmonary function (FEV1/Pred, DLCO/Pred, and SpO2 at rest and with exercise), will be listed by the patient.

7.7. Pharmacokinetic Analyses

The PK analyses will include patients with sufficient data to enable estimation of key parameters, and the parameters such as peak serum concentration ($C_{max,}$ postdose, the end of the infusion) and trough serum concentration ($C_{min,}$ predose) may be derived and summarized by visit/cycle at which these concentrations are collected with descriptive statistics such as mean, standard deviation, and coefficient of variation, etc.

Individual and/or mean serum ociperlimab and tislelizumab concentration versus time data will be tabulated and plotted as appropriate. Additional PK analyses may be conducted as appropriate.

8. INTERIM ANALYSIS (IA)

Two IAs are planned. The first interim analysis for non-binding futility will be conducted when approximately 245 deaths (65% of total target deaths) have been observed, which will occur approximately 37 months after the first patient is randomized.

An administrative alpha=0.0001 is spent at the first interim analysis. The second IA for efficacy is planned using the Lan-DeMets approach to the O'Brien-Fleming spending function using the remaining alpha=0.0249. It will be conducted after approximately 303 deaths (80% of total target deaths) have been observed, which will occur approximately 44 months after the first patient is randomized. The final analysis will be performed after approximately 379 death events have been observed, which will occur approximately 58 months after the first patient is randomized. Results from the two IAs will be reviewed by an IDMC.

Efficacy and futility stopping boundaries in p-value and Z score are shown in Table 4. The boundaries will be updated according to the actual numbers of events in the interim and final analyses, using the above pre-specified spending functions.

Table 4: Efficacy Stopping Boundaries (in p-value and Z score) of Primary Analyses of OS

Analysis	Time (months)	Number of events	p-Value [a] (Z score) for efficacy	p-Value [a] (Z score) for futility	Approximate HR threshold for efficacy	Approximate HR threshold for futility	Cumulative probability of crossing under H_1
Interim analysis 1 (IA1) *	37	245	0.0001 (3.719)	0.5294 (-0.074)	0.619	1.010	0.175
Interim analysis 2 (IA2)	44	303	0.0123 (2.248)		0.771		0.803
Final analysis (FA)	58	379	0.0214 (2.026)		0.811		0.930

Abbreviations: FA, final analysis; H_1, alternate hypothesis; HR, hazard ratio; IA1, first interim analysis; IA2, second interim analysis; mo, month; OS, overall survival.

[a] 1-sided, futility is non-binding

* An administrative alpha=0.0001 will be spent at the first interim analysis. However, there is no plan to stop the study for efficacy at this interim analysis.

IA will be performed by an independent statistician external to the Sponsor. The independent statistician will work with the blinded study statistician to provide statistical outputs to the IDMC as described in the IDMC charter and perform any ad-hoc analyses requested by the IDMC. If convincing evidence of OS is observed at the IA in the ITT population (i.e. the futility boundary at IA1 or efficacy boundary at IA2 is crossed), the IDMC will make a recommendation with regard to un-blinding/stopping the study early based on pre-defined criteria. Additional futility criteria are specified in the DMC Charter.

9. CHANGES IN THE PLANNED ANALYSIS

If the SAP needs to be revised, the sponsor will determine how the revision impacts the study and how the revision should be implemented. The details of the revision will be documented and described in the clinical study report.

10. REFERENCES

Brookmeyer R, Crowley J. A confidence interval for the median survival time. Biometrics. 1982;38:29-41.

National Cancer Institute Common Terminology Criteria for Adverse Events (NCI-CTCAE). Version 5.0. United States Department of Health and Human Services, National Institutes of Health, National Cancer Institute, Washington, DC, USA, November 27, 2017.

Eisenhauer EA, Therasse P, Bogaerts J Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eur J Cancer. 2009;45:228-47.

Greenwood M. The Natural Duration of Cancer. Reports on Public health and medical subjects. 1926;33:1-26.

Guidance for Industry: Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. Washington, DC: United States Food and Drug Administration; 2007.

ICH.Addendum on Estimands and Sensitivity Analyses in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials: E9(R1). 2019.

Mantel N, Haenszel W. Statistical Aspects of Analysis of Data from Retrospective Studies of Disease. Journal of the National Cancer Institute. 1959;22:719–748.

Robins JM, Tsiatis A. Correcting for non-compliance in randomized trials using rank-preserving structural failure time models. Communications in Statistics. 1991;20: 2609-31.

Sato T. On the Variance Estimator of the Mantel-Haenszel Risk Difference. Biometrics. 1989; 45:1323–1324.

APPENDIX 1. IMPUTATION OF MISSING OR PARTIALLY MISSING DATES

Please note:
The last known alive date is only based on complete dates without imputation. In general, all the imputed start date should be prior to/or by last known alive date.

1. Impute partial dates for concomitant medication

When the start date or end date of a medication/therapy/procedure is partially missing, the date will be imputed to determine whether the medication/therapy/procedure is prior or concomitant. The following rules will be applied to impute partial dates for medications.

If end date of a medication/therapy/procedure is partially missing, impute as follows:

- If both month and day are missing, then set to December 31
- If only day is missing, then set to last day of the month
- If the imputed end date > min(death date, study discontinuation date), then set to min (death date, study discontinuation date)

If start date of a medication/therapy/procedure is partially missing, impute as follows:

- If both month and day are missing, then set to January 01
- If only day is missing, then set to the first of the month
- If the imputed start date > min (death date, cutoff date, concomitant medication end date), then set to min (death date, cutoff date, concomitant medication end date).

If the year of start date or year of end date of a medication/therapy/procedure is missing, or the start date or end date is completely missing, do not impute.

2. Impute partial dates for adverse events

If year of the start date is missing or start date is completely missing, do not impute. Impute AE end date first if both AE start date and end date are partially missing.

If end date of an adverse event is partially missing, impute as follows:

- If both month and day are missing, then set to December 31
- If only day is missing, then set to last day of the month
- If the imputed end date > min(death date, study discontinuation date), then set to min (death date, study discontinuation date)

If year of the end date is missing or end date is completely missing, do not impute.

If start date of an adverse event is partially missing, impute as follows:

- If both month and day are missing and year = year of treatment start date, then set to treatment start date
- If both month and day are missing and year ≠ year of treatment start date, then set to

January 01

- If day is missing and month and year = month and year of treatment start date, then set to treatment start date
- If day is missing and month and year ≠ month and year of treatment start date, then set to first of the month
- If the imputed AE start date is after AE end date (maybe imputed), then update AE start date

with AE end date as final imputed AE start date

3. Impute partial dates related to disease history and prior therapy (Drug, surgery/procedure, radiotherapy)

The following rules will be applied to impute partial dates such as initial diagnosis date, initial BCLC staging date, relapse date, therapy date (start/end date), or surgery date etc.

Impute end date first. If end date is partially missing, impute as follows:

- If both month and day are missing, then set to December 31
- If only day is missing, then set to the last day of the month
- For prior systemic therapy for cancer, if imputed end date > randomization date – 6 months, then set to randomization date – 6 months *(in the event that 6 months is required per protocol)*
- For prior radiotherapy/locoregional therapy, if imputed end date > randomization date, then set to randomization date - 1

If start date is partially missing, impute as follows:

- If both month and day are missing, then set to January 01
- If only day is missing, then set to the first of the month
- If the imputed start date > end date, then set to the end date

If the year of start date or year of end date of a medication/therapy/procedure is missing, or the start date or end date is completely missing, do not impute.

4. Impute partial dates for subsequent anti-cancer therapy as collected in the post-treatment page (same rule applies to safety and efficacy flag)

If start date of subsequent anti-cancer therapy is partially missing, impute as follows:

- If both month and day are missing, then set to December 31
- If only day is missing, then set to last day of the month
- If the imputed start date > min (death date, study discontinuation date, data cutoff date, end date of subsequent anti-cancer therapy, start/end date of the next subsequent anti-cancer therapy), then set to min (death date, study discontinuation date, data

cutoff date, end date of subsequent anti-cancer therapy, start/end date of the next subsequent therapy)

If stop date of is partially missing, impute as follows:

- If both month and day are missing, then set to December 31
- If only day is missing, then set to last day of the month
- If the imputed stop date > min (death date, study discontinuation date, start date of the next subsequent anti-cancer therapy), then set to min (death date, study discontinuation date, start date of the next subsequent therapy)

The (imputed) stop date must be after or equal to the (imputed)start date

If year of the start date/stop date is missing, do not impute.

Note: if the imputed subsequent anti-cancer therapy date collected from CRF “post-treatment discontinuation anti-cancer systemic therapy” or “post-treatment discontinuation anti-cancer procedure” page is before the last dosing date, send data query.

APPENDIX 2. RULES FOR IDENTIFYING MISSING TUMOR ASSESSMENTS

Identifying two missing tumor assessments (TA)

1) Input scheduled TA visit list for each study
    a. 9wk-18wk-27wk-36wk-45wk-54wk-66wk- 78wk…
2) Identify last evaluable TA before PD or death (--LPTADT) and map it to the closest scheduled visit (--LPTADT_WK).
    a. In the event of unscheduled TA, choose the closest scheduled visit number (e.g. 9wk or 27wk) as –LPTADT_WK. It can be achieved programmatically by following the classification rule (e.g. defining thresholds) depicted in Table 5 below
    b. Otherwise, assign the scheduled visit number (assuming it is coded correctly) to --LPTADT_WK
3) Find the 2nd TA visit after LPTADT_WK according to the list in step 1 (--LPTADT_WK_2)
    a. If LPTADT_WK_2+1wk < earliest of PD/death date, then censor PFS at the –LPTADT

Table 5 shows how to assign unscheduled TA to a schedule visit. The Threshold column is defined as the mid-point between current and next visit (except for baseline); it is the upper limit for LPTADT to be mapped to the prior scheduled assessment (step 2a above). For example, if LPTADT is Week 47 for an unscheduled visit, it will be mapped to Week 45 TA since it is within the Threshold for Week 45. Assuming it is SD and the subsequent TA of the patient is PD after Week 67, PFS will be censored at LPTADT (Week 47); had the PD occurred prior to Week 67, it would be counted as an PFS event.

Table 5: Example of Scheduled Tumor Assessments with Time Window

Weeks	Scheduled week-1	**Scheduled week**	Scheduled week+1	Threshold
Baseline		Baseline		
Every 9 weeks for the first 52 weeks	Week 8	Week 9	Week 10	Week 13
	Week 17	Week 18	Week 19	Week 22
	Week 26	Week 27	Week 28	Week 31
	Week 35	Week 36	Week 37	Week 40
	Week 44	Week 45	Week 46	Week 49
Every 12 weeks afterwards	Week 53	**Week 54**	Week 55	Week 60
	Week 65	Week 66	Week 67	Week 72
	Week 77	Week 78	Week 79	Week 84
	Week 89	Week 90	Week 91	…