STATISTICAL ANALYSIS PLAN

Study Protocol Number:	BGB-A317-A1217-203 (AdvanTIG-203)
Study Protocol Title:	A Phase 2, Multicenter, Randomized, Placebo-Controlled Study to Compare the Efficacy of Anti-PD-1 Monoclonal Antibody Tislelizumab (BGB-A317) Plus Anti-TIGIT Monoclonal Antibody Ociperlimab (BGB-A1217) Versus Tislelizumab Plus Placebo as Second-Line Treatment in Patients With PD-L1 Tumor Area Positivity (TAP) ⩾ 10% Unresectable, Locally Advanced, Recurrent or Metastatic Esophageal Squamous Cell Carcinoma
Date:	27 February 2023
Version:	1.0
NCT Number:	NCT04732494

SIGNATURE PAGE

Author: [redacted] **Senior Statistician,** **Global Statistics and Data Science**	[redacted]

Approval

[redacted] **Vice President,** **Head of Global Statistics**	[redacted]
[redacted] **Executive Director,** **Clinical Development**	[redacted] Date: _____
[redacted] **Associate Medical Director,** **Clinical Development**	[redacted] _ Date: _____

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Definition
ADA	Antidrug Antibody
AE	Adverse Event
ATC	Anatomical Therapeutic Chemical
CBR	Clinical Benefit Rate
CR	Complete Response
CV	Coefficient of Variation
DCR	Disease Control Rate
DOR	Duration of Response
ECG	Electrocardiograms
ECOG PS	Eastern Cooperative Oncology Group Performance Status
eCRF	Electronic case report form
EORTC	European Organization for Research and Treatment of Cancer
EQ-5D-5L	European Quality of Life 5-Dimensional 5-Level
ESCC	Esophageal Squamous Cell Carcinoma
GHS	Global Health Status
HRQoL	Health Related Quality of Life
imAE	Immune-mediated adverse event
IRC	Independent Review Committee
INV	Investigator
IRR	Infusion-Related Reactions
IRT	Interactive Response Technology
ITT	Intent to Treat
MedDRA	Medical Dictionary for Regulatory Activities
NCI-CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Event
ORR	Objective Response Rate
OS	Overall Survival
PD	Progressive disease
PD-L1	Programmed cell death protein-ligand 1
PFS	Progression-free survival

PK	Pharmacokinetic
PR	Partial Response
PRO	Patient-reported outcomes
PT	Preferred Term
QLQ-C30	Quality of Life Questionnaire-Core 30
QLQ-OES18	Quality of Life Oesophageal Cancer Questionnaires 18
QoL	Quality of Life
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SI	Système International
SOC	System Organ Class
TAP	Tumor Area Positivity
TCM	Traditional Chinese Medication
TEAE	Treatment-Emergent Adverse Event
TMB	Tumor Mutation Burden
TTR	Time to Response
VAS	Visual Analog Scale
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 AdvanTIG-203: a Phase 2, multicenter, randomized, placebo-controlled study to compare the efficacy of anti-PD-1 monoclonal antibody tislelizumab (BGB-A317) plus anti-TIGIT monoclonal antibody ociperlimab (BGB-A1217) versus tislelizumab plus placebo as second-line treatment in patients with PD-L1 TAP ≥ 10% unresectable, locally advanced, recurrent or metastatic esophageal squamous cell carcinoma. This SAP is based on AdvanTIG-203 Protocol Amendment 2.0, dated on 27 February, 2023. The focus of this SAP is the planned final analysis specified in the study protocol. Separate analysis plans for Pharmacogenomics and Biomarker analyses that are not described within this SAP may be prepared as needed.

2. STUDY OVERVIEW

2.1. Study Design

This is a multicenter, randomized, investigator-and patient-blinded, sponsor-unblinded, placebo-controlled global Phase 2 study to compare the efficacy of anti-PD-1 monoclonal antibody tislelizumab plus anti-TIGIT monoclonal antibody ociperlimab versus tislelizumab plus placebo as second-line treatment in patients with PD-L1 TAP ≥ 10% unresectable, locally advanced, recurrent or metastatic Esophageal Squamous Cell Carcinoma (ESCC). After providing written informed consent, completing all prescreening and screening assessments, and being confirmed as eligible for study participation, approximately 120 patients will be randomized at a 1:1 ratio to receive 1 of the following treatment regimens:

- Arm A: tislelizumab + ociperlimab
- Arm B: tislelizumab + placebo

Eligible patients will be stratified by the following factors:

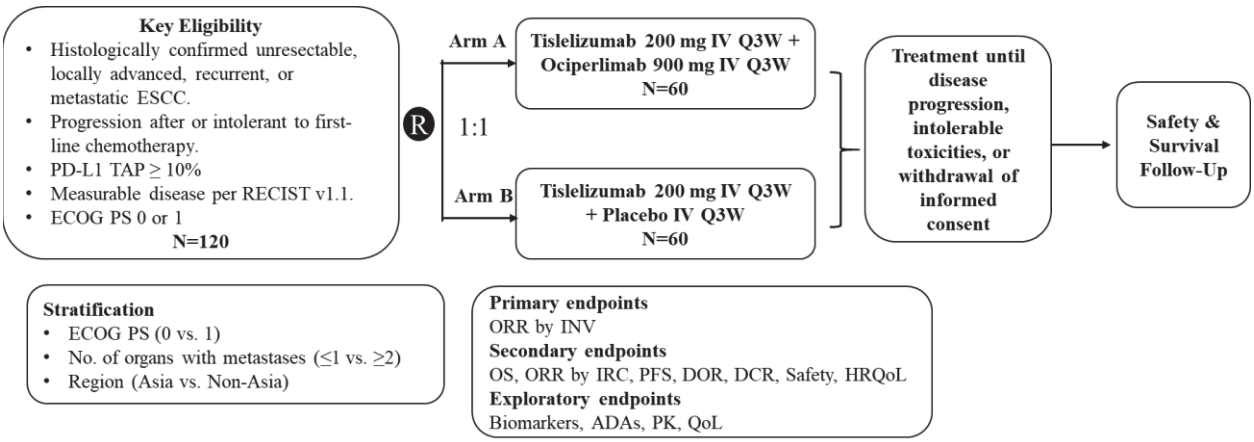
- ECOG PS score (0 versus 1)
- Number of organs with metastases (≤1 versus ≥2)
- Region (Asia versus non-Asia)

Study drugs (including placebo) will be administered until PD per RECIST v1.1, unacceptable toxicity, or withdrawal of informed consent, whichever occurs first.

No crossover between Arm A and Arm B will be allowed.

The study design schema is in Figure 1.

Figure 1: Study Schema



Abbreviations: ADA, antidrug antibody; DCR, disease control rate; DOR, duration of response; ECOG PS, Eastern Cooperative Oncology Group Performance Status; ESCC, esophageal squamous cell carcinoma; HRQoL, Health Related Quality of Life; INV, investigator; IRC, Independent Review Committee; IV, intravenously; ORR, objective response rate; OS, overall survival; PD-L1, programmed cell death protein-ligand 1 (PD-L1); PFS, progression-free survival; PK, pharmacokinetics; Q3W, once every 3 weeks; RECIST v1.1, Response Evaluation Criteria in Solid Tumors version 1.1; TAP; Tumor Area Positivity.

Note: The initial infusion (Cycle 1 and Cycle 2, Day 1) will be administered over a period of 60 minutes. If this infusion is well tolerated, subsequent infusions may be administered over 30 minutes. After infusion, patients will be further monitored for at least 1 hour during Cycles 1 and 2. From Cycle 3 onward, a postinfusion monitoring period of at least 30 minutes will be required.
At the discretion of the investigator, patients may continue tislelizumab and ociperlimab or tislelizumab alone after PD under protocol defined conditions.

Key changes made from Original Protocol 0.0 (16 September 2020) to Protocol Amendment 1.0 (03 June 2022) were as follows:

- The expected number of randomized patients was revised from approximately 280 to approximately 120.
- Overall survival (OS) was moved from a primary objective to secondary objective.
- Study became single-blind versus double-blind (sponsor is unblinded but investigators, site staff, and patients remains blinded).

Key changes made from Protocol Amendment 1.0 (03 June 2022) to Protocol Amendment 2.0 (27 February 2023) were as follows:

- Revised the percentage and number of the overall survival (OS) events from 70% (84 deaths) of the total sample size to 60% (72 deaths) in the secondary efficacy analysis.
- Revised the testing method of the null hypothesis of ORR from Miettinen and Nurminen test to Cochran-Mantel-Haenszel method.

2.2. Study Assessments

Tumor Assessment

Tumor imaging will be performed at baseline (≤ 28 days before randomization). During the study, tumor imaging will be performed approximately every 6 weeks (± 7 days) for the first 54 weeks and every 12 weeks (± 7 days) thereafter. Response will be assessed using RECIST v1.1. A patient who discontinues study drugs early for reasons other than PD (eg, treatment toxicity) will continue to undergo tumor assessments according to the original plan until the patient experiences PD, withdraws consent, is lost to follow-up, dies, or until the study terminates, whichever occurs first.

Patient-reported outcomes (PROs) will be collected using the EORTC QLQ-C30, EORTC QLQ-OES18, and EQ-5D-5L before dosing on Day 1 of every other treatment cycle during the first 54 weeks, Day 1 of every 4 cycles thereafter, and at the on-site safety follow-up visit.

Safety Assessment

Patients will be evaluated for any AEs and SAEs occurring up to 30 days after the last dose of study drugs (all severity grades, per NCI-CTCAE v5.0) or initiation of a new anticancer therapy, whichever occurs first, and immune-mediated AEs (imAEs) occurring up to 90 days after the last dose of study drugs regardless of initiation of a subsequent anticancer therapy. All drug-related SAEs will be recorded by the investigator after treatment discontinuation until patient death, withdrawal of consent, or loss to follow-up, whichever occurs first, and will be followed until they resolve to baseline or ≤ Grade 1, are assessed by the investigator as stable and unlikely to improve, or the patient is lost to follow-up, whichever occurs first.

3. STUDY OBJECTIVES

3.1. Primary Objective

- To compare the objective response rate (ORR), as assessed by the investigator according to RECIST v1.1, of tislelizumab plus ociperlimab with tislelizumab plus placebo as second-line treatment in patients with PD-L1 TAP ≥ 10% unresectable, locally advanced, recurrent or metastatic ESCC in the Intent to Treat (ITT) Analysis Set.

3.2. Secondary Objective

- To compare the overall survival (OS) of tislelizumab plus ociperlimab with tislelizumab plus placebo as second-line treatment in patients with PD-L1 TAP ≥ 10% ESCC in the ITT Analysis Set.
- To compare the following endpoints between tislelizumab plus ociperlimab and tislelizumab plus placebo based on tumor assessments per RECIST v1.1:
  - Progression-free survival (PFS) assessed by both Independent Review Committee (IRC) and the investigator
  - Duration of response (DOR), DCR, and clinical benefit rate (CBR) assessed by both the IRC and the investigator
  - ORR assessed by the IRC

- To compare the safety and tolerability between tislelizumab plus ociperlimab and tislelizumab plus placebo.
- To compare the health-related quality of life (HRQoL) via cancer-specific patient reported outcomes (PROs) between tislelizumab plus ociperlimab and tislelizumab plus placebo.

3.3. Exploratory Objective

- To characterize the PK of ociperlimab and tislelizumab.
- To assess host immunogenicity to ociperlimab and tislelizumab.
- To compare the quality of life (QoL) via a generic PRO between tislelizumab plus ociperlimab and tislelizumab plus placebo.

4. DEFINITION OF ESTIMANDS

Primary scientific question of interest: "Will tislelizumab plus ociperlimab versus tislelizumab plus placebo increases complete response (CR) or partial response (PR) rate prior to disease progression and any new anticancer therapy in patients with PD-L1 TAP ≥ 10% unresectable, locally advanced, recurrent or metastatic ESCC?"

The primary estimand is described by the following attributes:

1. Treatment of interest
   The study experimental treatment is tislelizumab + ociperlimab. The study control treatment for comparison is tislelizumab + placebo.
2. Population
   Adult patients with PD-L1 TAP ≥ 10% unresectable, locally advanced, recurrent or metastatic ESCC who progressed or were intolerant to first-line treatment.
3. Variable
   ORR, defined as the proportion of patients who have confirmed CR or PR prior to disease progression by the investigator's review per RECIST v1.1.
4. Handling of remaining intercurrent events
   - New anticancer therapy started prior to disease progression: patients starting any new anticancer therapy without achieving a confirmed CR or PR before will be considered as non-responders (composite strategy).
   - Discontinuation of treatment prior to disease progression: response assessment after discontinuation of treatment will be counted and used for analysis (treatment policy strategy)
5. Population-level summary
   The ORR difference estimated by Mantel-Haenszel common risk difference stratified by, ECOG PS score [0 versus 1] and the number of organs with metastases [≤1 versus ≥2]).

5. STUDY ENDPOINTS

5.1. Primary Endpoint(s)

- ORR, defined as the proportion of patients who have confirmed CR or PR by the investigator's review per RECIST v1.1 in the ITT Analysis Set.

5.2. Secondary Endpoints

- OS, defined as the time from the date of randomization until the date of death due to any cause in all randomized patients in the ITT Analysis Set.
- ORR, defined as above and assessed by the IRC per RECIST v1.1 in the ITT Analysis Set.
- PFS, defined as the time from the date of randomization to the date of first documentation of PD assessed by investigator (or IRC) per RECIST v1.1 or death, whichever occurs first in the ITT Analysis Set.
- DOR, defined as the time from the first determination of an objective response until the first documentation of PD as assessed by investigator (or IRC) per RECIST v1.1, or death, whichever comes first in the ITT Analysis Set.
- DCR, defined as the proportion of patients who have confirmed CR, confirmed PR, and stable disease assessed by investigator (or IRC) per RECIST v1.1 in the ITT Analysis Set.
- CBR, defined as the proportion of patients who achieve confirmed CR, confirmed PR, and durable stable disease (stable disease ≥ 24 weeks).
- HRQoL, assessed by changes from baseline in the scores of key PRO endpoints of Global Health Status (GHS)/QoL and Physical Function of European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire-Core 30 (QLQ-C30) and scores of Dysphagia, Eating, Reflux and Pain scales of EORTC Quality of Life Oesophageal Cancer Questionnaires 18 (QLQ-OES18).
- AEs and SAEs as characterized by type, frequency, severity (as graded by National Cancer Institute Common Terminology Criteria for Adverse Event [NCI-CTCAE] version 5.0 [v5.0]), timing, seriousness, and relationship to study drugs, physical examinations, electrocardiograms (ECGs), and laboratory assessments.

5.3. Exploratory Endpoints

- Serum ociperlimab and tislelizumab concentrations at specified timepoints.
- Immunogenic responses to ociperlimab and tislelizumab, evaluated through the detection of ADAs.
- QoL, defined as assessment of the European Quality of Life 5-Dimensional 5-Level (EQ-5D-5L) via visual analog scale (VAS) in the ITT Analysis Set.

6. SAMPLE SIZE CONSIDERATIONS

The sample size calculation is based on the primary efficacy analysis of ORR between Arm A and Arm B in the ITT Analysis Set. The number of patients was revised from 280 to 120 based on the Protocol Amendment 1.0.

With 120 randomized patients, the study has at least 80% power to detect a 24% difference in ORR (45% vs 21% in Arm A and Arm B, respectively) at a 1-sided type I error of 0.025.The efficacy boundary of primary analysis of ORR is >16.3% ORR difference.

The sample size is calculated by EAST (version 6.4.1).

7. STATISTICAL METHODS

7.1. Analysis Sets

The Intent-to-Treat (ITT) Analysis Set includes all randomized patients. Patients will be analyzed according to their assigned treatment at randomization. This will be the primary analysis set for all efficacy analyses, including HRQoL analysis.

The Safety Analysis Set includes all patients who received ≥ 1 dose of any component of study drugs. Patients will be analyzed according to their actual treatment received. This will be the primary analysis set for safety analyses.

The PK Analysis Set includes all patients who received ≥ 1 dose of any component of study drugs per the protocol, and for whom any postdose PK data are available.

The Immunogenicity Analysis Set includes all patients who received ≥ 1 dose of any component of study drugs and for whom both baseline antidrug antibody and at least 1 postbaseline antidrug antibody result are available.

7.2. Multiplicity Adjustment

The primary endpoint ORR will be tested at a 1-sided alpha of 0.025. If the null hypothesis for ORR in the ITT Analysis Set is rejected, the secondary endpoint OS in the ITT Analysis Set will be tested.

7.3. Data Analysis General Considerations

7.3.1. Definitions and Computations

Study day

Study day will be calculated in reference to the first dose date for safety analysis. For assessments conducted on or after the first dose date, study day will be calculated as (assessment date –first dose date + 1). For assessments conducted before first dose date, study day is calculated as (assessment date – first dose date). There is no study day 0. In the situation where the event date is partial or missing, the date will appear partial or missing 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 date of randomization.

Baseline Measurements

- Efficacy evaluation except HRQoL

  A baseline value is defined as the last non-missing value collected prior to the randomization.

- Safety and HRQoL

  A baseline value is defined as the last non-missing value prior to the first study drug administration.

- 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 LLN, two baseline toxicity grades: Grade 1 for Hypo and Grade 0 for Hyper will be derived.

Study Follow-up Duration

Study follow-up duration is defined as the duration from the randomization date to the study discontinuation date (eg, death, consent withdrawal, lost to follow-up) or to cutoff date if a patient is still ongoing.

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.
- 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 radiograph was obtained rather than the associated visit date.
- For laboratory results presented in numerical range, if lab results >= x or > x, then set it as x; if lab results <= x or < x, then set it as x/2.
- For by-visit observed data analyses, percentages will be calculated based on the N unless it's specified in the footnote.
- 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

Handling of missing data related to primary estimand will be further elaborated 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. Other missing data will not be imputed unless otherwise specified elsewhere in this SAP.

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 randomized, treated, discontinued from the study, reasons for discontinued from the study, and the duration of study follow-up will be summarized in the ITT analysis set. The patients who discontinued treatment and the primary reason for the end of treatment will be summarized among patients who were treated. The reasons for treatment/study discontinuation related to COVID-19 impact will also be summarized.

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. Patients with multiple important protocol deviations in each category/subcategory were counted only once at the category/subcategory level.

Critical protocol deviation that significantly impacts efficacy or safety evaluation will be reviewed prior to database lock according to the criteria defined in protocol deviation specification.

7.4.3. Demographic and Other Baseline Characteristics

Demographics 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 (Male, Female)
- Race
- Ethnicity
- Country
- Baseline Weight (kg)
- Baseline BMI (kg/m^2)

In addition, the stratification factors per Interactive Response Technology (IRT) and per eCRF will be summarized using ITT analysis set:

- ECOG PS score (0 versus 1)
- Number of organs with metastases (≤1 versus ≥2)
- Region (Asia versus non-Asia)

7.4.4. Disease History

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

- Stage at Initial Diagnosis (Stage I, II, III, IV, unknown)
- Time from Initial Diagnosis to Study Entry (month)
- Primary Site of Esophageal Cancer (cervical, upper, middle, and lower)
- Histologic Grade (Gx, G1, G2, G3, not done)
- Target lesions sum of diameter by investigator (mm)
- Metastatic Disease Status at Study Entry (yes, no)
- Time from Date of Metastatic Diagnosis to Study Entry (month)
- Anatomic locations of metastases
- Previous treatment (chemotherapy, radiation therapy, and surgery)

7.4.5. Prior Anticancer Drug Therapies and Surgeries

Prior anticancer drug therapies, prior anticancer radiotherapy, and prior anticancer surgeries will be summarized in the ITT analysis set. For prior anticancer drug therapies, the variables include number of patients with any prior anticancer therapy, number of prior lines, treatment setting, duration of prior anticancer therapy, reasons for discontinuation of last anticancer drug therapy, best overall response of the last prior anticancer drug therapy, time from the end of the last prior anticancer therapy to randomization, and time from last disease progression to randomization. For prior anticancer radiotherapy, the variables include number of patients with any prior radiotherapy, treatment intent, treatment setting, sites irradiated, and time from end of last radiotherapy to randomization. For prior anticancer surgeries, the variables include number of patients with any prior anticancer surgery, time from last anticancer surgery to randomization, and prior anticancer surgery location . 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 will be defined as medications that stopped before the day of the first dose of study drugs. Concomitant medications will be defined as medications that 1) started before the first dose of study drugs and were continuing at the time of the first dose of study drugs, or 2) started on or after the date of the first dose of study drugs up to 30 days after the patient's last dose.

Prior and concomitant medications will be coded using the World Health Organization Drug Dictionary (WHO DD) drug codes Version B3 March 2022 or higher and 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 safety analysis set.

7.4.7. Medical History

Medical history will be coded using the latest version of Medical Dictionary for Regulatory Activities (MedDRA) terminology available at the time of the analyses. 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 in the ITT analysis set. A listing of medical history will be provided.

7.5. Efficacy Analysis

7.5.1. Primary Efficacy Endpoint(s)

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

Variable

ORR is defined as the proportion of patients who have confirmed CR or PR by the investigator per RECIST v1.1 in the ITT Analysis Set. Patients with no postbaseline response assessment (for any reason) will be considered non-responders.

Primary Analysis for Primary Estimand

The null hypothesis to be tested is:

$$H_0: \text{ORR in Arm A} \leqslant \text{ORR in Arm B}$$

against the alternative:

$$H_1: \text{ORR in Arm A} > \text{ORR in Arm B}$$

The null hypothesis will be tested using the Cochran-Mantel-Haenszel method stratified by the selected stratification factors at randomization (ie, ECOG PS score [0 versus 1] and the number of organs with metastases [$\leqslant$1 versus $\geqslant$2]) at the one-sided significance level of 0.025. If the null hypothesis can be rejected, it will be concluded that the ORR in Arm A (Tislelizumab plus BGB-A1217) is superior to the ORR in Arm B (Tislelizumab plus placebo).

A Mantel-Haenszel common risk difference (Mantel and Haenszel, 1959) will be estimated, along with its 95% confidence interval constructed by a normal approximation and Sato's variance estimator (Sato, 1989) with the same stratification factors as described above. The Clopper Pearson 95% CIs for the ORR within each arm will also be calculated.

The final analysis of ORR is planned approximately 4 months after the randomization of the last patient.

Further derivation rules for the primary analysis of ORR are presented in Table 1.

Table 1: Handling of Intercurrent Events and Missing Values for the Primary Analysis of ORR

	Derivation rules
Intercurrent events	
New anticancer therapy started prior to disease progression	Patients starting any new anticancer therapy without achieving a confirmed CR or PR before will be considered as non-responders (composite strategy)
Discontinuation of treatment prior to disease progression	Response assessment after discontinuation of treatment will be counted and used for analysis (treatment policy strategy)
Missing values not related to intercurrent events	
No post-baseline response assessment (regardless of the reason)	Non-responders

Sensitivity analyses

Sensitivity analysis 1

ORR assessed by the investigator's review will be analyzed using an unstratified Cochran-Mantel-Haenszel method. The crude rate difference along with its 95% CI will be provided. This analysis relaxes the assumption made by the primary analysis for the primary estimand that the baseline probabilities of achieving PR/CR is different among the strata (i.e., the stratification factors are confounding).

Sensitivity analysis 2

ORR assessed by the investigator's review will be analyzed using a Cochran-Mantel-Haenszel method stratified by the actual value of selected stratification factors collected from eCRF (ie, ECOG PS score [0 versus 1] and the number of organs with metastases [≤1 versus ≥2]).

Supplementary analysis

This analysis targets an estimand which has the same attributes as the primary estimand except that the variable changed to unconfirmed ORR assessed by the investigator's review. The same analysis methods as the one for the primary estimand will be used.

7.5.2. Secondary Efficacy Endpoints

Overall Survival

Overall survival is defined as time from 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. The last known alive date 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, if day of death date is missing, death date is imputed as the max (last available date showing patient was alive + 1, first day of year/month of death date). The patient with imputed death date will be considered as an event for OS analysis.

Primary analysis of Overall Survival

The null hypothesis to be tested is:

H_0: OS in Arm A ≤ OS in Arm B

against the alternative:

H_1: OS in Arm A > OS in Arm B

The null hypothesis will be tested using a log-rank test stratified by the selected stratification factors at randomization (ie, ECOG PS score [0 versus 1] and the number of organs with metastases [≤1 versus ≥2]).

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 OS, including median, Q1 and Q3, and event-free rates at every 3 months, will be estimated using the Kaplan-Meier method for each treatment group. 95% 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). Kaplan-Meier survival probabilities over time for each arm will be plotted.

The primary analysis of OS will take place when approximately 72 deaths (ie, 60% of the total sample size) have been observed.

Sensitivity analysis

Sensitivity analysis 1

OS will be analyzed using a unstratified Cox model, and the treatment effect will be summarized by the hazard ratio with its 95% confidence interval.

Sensitivity analysis 2

OS will be analyzed using stratified Cox model and the distribution of OS will be compared between the treatment group using stratified log-rank test, with stratification factors obtained from the eCRF.

Overall response rate by IRC

ORR assessed by the IRC per RECIST v1.1 in the ITT Analysis Set will be analyzed similarly as the primary analysis of ORR assessed by the investigator. The Mantel-Haenszel common risk difference stratified by the selected stratification factors at randomization (ie, ECOG PS score [0 versus 1] and the number of organs with metastases [≤1 versus ≥2]), along with its 95% confidence interval constructed by a normal approximation and Sato's variance estimator (Sato, 1989) will be estimated. The Clopper Pearson 95% CIs for the ORR within each arm will also be calculated. The Cochran-Mantel-Haenszel method stratified by the selected stratification factors at randomization (ie, ECOG PS score and the number of organs with metastases [≤1 versus ≥2]) will

be provided to show the magnitude of ORR difference, and are only used for descriptive purpose only. Unconfirmed ORR assessed by the IRC will also be analyzed.

Progression Free Survival

PFS is defined as the time from the date of randomization to the date of first documentation of PD assessed by investigator (or IRC) per RECIST v1.1 or death, whichever occurs first in the ITT Analysis Set. Progression-free survival censoring rule in handling intercurrent events and missing values are described in Table 2 , which follows United States (US) Food and Drug Administration (FDA) Guidance for Industry, Clinical Trial Endpoints for Approval of Cancer drugs and Biologics (2018). The algorithm to identify missing TAs are presented in Appendix 2.

Table 2: Censoring Rules for Progression-Free Survival Per RECIST Version 1.1

	Derivation rules	**Outcome**
No progression at the time of data cut-off or withdrawal from study or lost to follow up	Date of last adequate radiologic assessment prior to or on date of data cut-off or withdrawal from study or lost to follow up	Censored
New anticancer therapy started prior to disease progression or death	Last adequate radiological assessment before the new anticancer therapy (hypothetical strategy)	Censored
No baseline or post-baseline tumor assessments without death within 13 weeks after randomization	Date of randomization	Censored
No baseline or post-baseline tumor assessments with death within 13 weeks after randomization	Date of death	Event
Death or progression after more than one missed visit	Date of last adequate radiologic assessment before missed tumor assessments	Censored

The distribution of PFS, including median, Q1, Q3, and event-free rates at every 3 months, will be estimated using the Kaplan-Meier method for each treatment group. 95% CIs for median, 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). The treatment effect will be estimated by fitting a Cox regression model with treatment arm adjusted by the before mentioned stratification factors as strata. From this model, the HR of PFS will be estimated and presented with a 2-sided 95% CI.

Duration of Response

Duration of Response (DOR) is defined as the time from the first determination of an objective response until the first documentation of PD as assessed by investigator (or IRC) per RECIST v1.1, or death, whichever comes first in the ITT Analysis Set. All the censoring rules for PFS will be applied to DOR too. DOR will be analyzed in the responders only. The distribution of DOR, including median, Q1, Q3, and event-free rates at every 3 months, will be estimated using the Kaplan-Meier method for each treatment group. 95% CIs for median, Q1, and Q3 of DOR 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). No formal testing will be performed to compare DOR between two treatment group as it would be based on a non-randomized subgroup. Both confirmed and unconfirmed DOR will be analyzed.

Disease control rate (DCR), clinical benefit rate (CBR)

DCR, defined as the proportion of patients who have confirmed CR, confirmed PR, and stable disease assessed by investigator (or IRC) per RECIST v1.1 in the ITT Analysis Set. CBR, defined as the proportion of patients who achieve confirmed CR, confirmed PR and durable stable disease (stable disease≥ 24 weeks) assessed by investigator (or IRC) per RECIST v1.1 in the ITT Analysis Set. DCR and CBR will be analyzed similarly as the primary analysis of ORR in the ITT Analysis Set. The Mantel-Haenszel common risk difference, along with its 2-sided 95% CIs will be calculated, as will the Clopper-Pearson 95% CIs for the DCR and CBR within each arm. Unconfirmed DCR and CBR may be analyzed if needed.

Waterfall plots will be provided for the maximum tumor shrinkage based on target lesions. The maximum tumor shrinkage based on target lesion used in the plots will be listed. These analyses will be performed based on RECIST1.1.

Time to response (TTR) will be summarized using descriptive statistics, such as mean, median, and standard deviation. Only patients who have achieved objective response by investigators or IRC per RECIST v1.1 will be included in the analysis of TTR.

Health -Related Quality of Life (HRQoL)

The EORTC-QLQ-C30 consists of thirty questions that are associated with one global health status/QoL (GHS) scale (Aaronson NK, et al., 1993; Fayers PM, et al., 2001), five functional scales (physical, role, cognitive, emotional, and social), three symptom scales (fatigue, pain, and nausea and vomiting), and six single item scales (dyspnoea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties).

The EORTC-QLQ-OES18 (Wen Y, et al., 2015) is the specific esophageal symptoms module of the QLQ-C30, and includes 18 questions: 6 single item subscales measuring saliva swallowing, choking, dry mouth, taste, coughing, and talking. It also includes 12 items grouped into 4 subscales: dysphagia (3 items), eating (4 items), reflux (2 items), and pain (3 items).

EORTC Scoring Derivation

The principle for scoring applies to all scales/scores: Raw scores are calculated as the average of the items that contribute to the scale. A linear transformation to standardize the raw scores is utilized, so that the scores are ranged from 0 to 100. Increases in scores for functional domains (e.g., physical, role, social, emotional, etc.) are improvements while increases in scores for symptoms (e.g., fatigue, vomiting and nausea, diarrhea, pain, etc.) are deteriorations.

If at least half of the items for a scale are answered, then all the completed items are used to calculate the scale score. Otherwise, the scale score is set to missing.

Raw Score (RS)

For all scores, the raw score (RS) is the mean of the component items.

$$RS = (I_1+I_2+\ldots+ I_n)/n$$

Derived Scale (DS)

The derived scales are obtained from the raw scores as defined in the EORTC manual. The derived scales have a more intuitive interpretation: larger function scale or global health status / QoL are improvements while larger symptom scales (e.g., pain, nausea, etc.) are deteriorations. The derivation formulas are as follows.

For functional scales:

$$DS = [1 - (RS-1)/range] \times 100$$

For symptom scales and global health status:

$$DS = [(RS-1)/ range] \times 100$$

Refer Table 3 and Table 4 for EORTC -QLQ-C30 and EORTC-QLQ-OES18 scoring.

Index Score

The domain scores and single item scores are derived scales. Below is the index score formula.

Index score = [∑(domain scores + single item scores)]÷ number of available domains and single items

Table 3: Scoring of QLQ-C30

	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				
Fatigue	FA	3	3	10, 12, 18
Nausea and vomiting	NV	2	3	14, 15
Pain	PA	2	3	9, 19
Single Items				
Dyspnoea	DY	1	3	8
Insomnia	SL	1	3	11

Appetite loss	AP	1	3	13
Constipation	CO	1	3	16
Diarrhea	DI	1	3	17
Financial Difficulties	FI	1	3	28

Table 4: Scoring of QLQ-OES18

	Scale	Number of items	Item range	Item Numbers
Symptom Scales				
Dysphagia	DY	3	3	31, 32, 33
Eating	EA	4	3	36, 37, 38, 39
Reflux	RE	2	3	44, 45
Pain	PA	3	3	46, 47, 48
Single Items				
Trouble swallowing saliva	SA	1	3	34
Choked when swallowing	SW	1	3	35
Dry mouth	DM	1	3	40
Trouble with taste	TA	1	3	41
Trouble with coughing	CO	1	3	42
Trouble talking	TA	1	3	43

All HRQoL measures will be summarized in ITT analysis set.

A questionnaire module is considered complete if at least one derived scale is answered. The adjusted completion rate is defined as percentages of patients who completed the questionnaire at each visit divided by the number of patients still in treatment who were expected to complete the questionnaire. Completion rates and adjusted completion rates for the EORTC-QLQ-C30 and EORTC-QLQ-OES18 will be summarized separately at each visit.

The derived scores (functional scales/symptom scales/single items and GHS/QoL scale) of EORTC-QLQ-C30 and EORTC-QLQ-OES18 and the index score of QLQ-OES18 will be summarized as well as change from baseline at each visit using descriptive statistics. The plots of the mean values and their standard errors at each visit over time for selected domains of EORTC-QLQ-C30 and EORTC-QLQ-OES18 by treatment arm will be presented for both Arm A and Arm B.

A mixed effect model analysis for assessing differences between the arms in changes post-baseline will be performed for the PRO endpoints (GHS/QoL scale and physical function measured

by QLQ-C30; and dysphagia, eating, reflux, pain measured by QLQ-OES18). Difference in change from baseline to the key clinical of cycle 5 and cycle 7 between treatment arms will be assessed. The model includes the PRO endpoint as dependent variable, 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. LS mean change differences between arms at 95% CI with nominal p values will be reported.

Time to deterioration analysis will be performed to compare the risk of experiencing a deterioration event in the PRO end points (GHS/QoL scale and physical function measured by QLQ-C30; and dysphagia, eating, reflux, pain measured by QLQ-OES18) between two treatment arms. Time to deterioration is defined as changes ≥ 10 points (Osaba et al 1998) from baseline in the worsening direction. A deterioration is not counted as an event if a subsequent improvement returned to less than 10 points. If a patient does not have an deterioration event, they are censored at their last clinic visit at which HRQoL score is measured. The log-rank test and hazard ratio will be provided to show the magnitude of treatment effect and are only used for descriptive purpose only.

7.5.3. Subgroup Analyses

Subgroup analysis on ORR by the investigator's review and OS will be conducted in ITT analysis set, to evaluate the uniformity of treatment effect across variety of subgroups. Subgroup variables may include, but not limited to,

- Number of metastatic lesions (≤ 1 site versus ≥ 2 sites)
- Geographic region: (Asia/Europe)
- ECOG Performance status: (0, 1)
- Age group: (< 65 years, >= 65 years)
- Sex: (Male, Female)
- Smoking status: (former/current smoker, non-smoker)
- Race: (White, Asian, and Other)
- Previous treatment (radiation therapy, surgery)

For ORR by the investigator's review, in each subgroup, a crude risk difference will be estimated, along with its 95% confidence interval . And the Clopper Pearson 95% CIs for the ORR within each arm will also be calculated.

For OS, KM estimates by treatment arm, and the estimation of unstratified hazard ratio and their 95% CI will be provided for the subgroups.

7.5.4. Exploratory Efficacy Endpoints

QoL

The EQ-5D-5L comprises a descriptive module and an EQ Visual Analogue scale (EQ VAS). The EQ-5D-5L descriptive module comprises the following 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems, and extreme problems. The EQ VAS

records the respondent's self-rated health on a 0 to 100 scale, with 100 labelled 'the best health you can imagine' and 0 'the worst health you can imagine'. Lower scores in descriptive dimension indicates better HRQoL and higher VAS score indicates better health state.

EQ-5D-5L descriptive module will be summarized by visit and dimension in an ordinal scale using descriptive statistics. The EQ VAS and change from baseline will be summarized by visit in a continuous scale using descriptive statistics. Completion rates for EQ-5D-5L will be summarized at each visit. A questionnaire module is considered complete if at least one question is answered.

7.5.5. Post and during-treatment Anticancer Therapy

New anticancer therapy

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

- As for efficacy analysis, the start date of new anticancer therapy will be the earliest date of prohibited anticancer therapy taken during treatment, date of the post-treatment systemic anticancer therapy and date of other anticancer therapy such as post-treatment surgery and radiotherapy as deemed appropriate.
- For non-efficacy analysis, the start date of new anticancer therapy is always the first date of new systemic anticancer 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 Traditional Chinese Medication (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 anticancer therapy in the efficacy and safety analyses.

Subsequent systemic anticancer therapy

Subsequent systemic anticancer therapy is defined as the anticancer therapy started after the last dose of study drug(s). A summary of number and percentage of patients who received subsequent systematic anticancer therapy by therapy type (e.g. chemotherapy, immunotherapy, targeted therapy) will be provided by treatment arm based on ITT analysis set.

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 AEs, laboratory values, vital signs, ECG findings and physical examination.

7.6.1. Extent of Exposure

The following measures of the extent of exposure will be summarized for ociperlimab/placebo and tislelizumab separately:

- Duration of exposure (days): defined as last date of exposure – first dose date + 1.
  - If patients discontinue treatment (with non-missing EOT date), last date of exposure = min (last dose date + 20, death date, clinical data cutoff date)

  - Otherwise, if patient has treatment ongoing, last date of exposure = clinical data cutoff date

- Number of treatment cycles received: defined as the total number of cycles with non-missing doses.
- Total dose received per patient (mg): defined as the sum of all actual dosages per administration at all visits prior to the clinical data cutoff date.
- Actual dose intensity (mg/cycle): defined as 21 × total dose received per patient (mg) / (last dose date prior to cutoff date + 21 – first dose date)
- Relative dose intensity (%): defined as the ratio of the actual dose intensity and the planned dose intensity. The planned dose intensity is 200 mg/cycle for tislelizumab and 900 mg/cycle for ociperlimab/placebo.

The number of patients with dose modifications which includes dose missed, dose delays and infusion interruptions and their reasons will be summarized by counts and percentages according to study drug.

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

7.6.2. Adverse Events

AEs will be graded by the investigators using CTCAE version 5.0. 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 25.0 or higher) lowest level term closest to the verbatim term, along with the linked MedDRA preferred term (PT) and primary system organ class (SOC).

7.6.2.1 Treatment Emergent Adverse Event

A treatment-emergent adverse event (TEAE) is defined as an AE that had an onset date or a worsening in severity from baseline (pretreatment) on or after the first dose of study drugs up to 30 days of study drugs after last dose or initiation of subsequent anticancer therapy, whichever comes first. Only those AEs that were treatment emergent will be included in summary tables of TEAE. All AEs, treatment emergent or otherwise, will be presented in patient data listings.

An AE overview table, including the number and percentage of patients with TEAEs, treatment-emergent serious adverse events (SAEs), TEAEs with Grade 3 or above, TEAEs that led to death, TEAEs that led to treatment discontinuation, TEAEs that led to dose delay, TEAEs that led to infusion interruption/rate decrease, treatment-related TEAEs, treatment-related version of any of the above categories, infusion-related reactions will be provided. Treatment-related AEs include those events considered by the investigator to be related to study drug or with a missing assessment of the causal relationship.

The incidence of TEAEs will be reported as the number (percentage) of patients with TEAEs by SOC and PT. A patient will be counted only once by the highest severity grade within an SOC and PT, even if the patient experienced more than 1 TEAE within a specific SOC and PT.

The number (percentage) of patients with following TEAE will be provided:

- All TEAEs
  - TEAEs by SOC and PT (any grades and >= grade 3)
  - TEAE by PT (any grades and >= grade 3)
  - Treatment-related TEAEs by SOC and PT (any grades and >= grade 3)
  - Treatment-related TEAEs by PT (any grades and >= grade 3)
  - Tislelizumab-related TEAEs by SOC and PT (any grades and >= grade 3)
  - Ociperlimab/Placebo-related TEAEs by SOC and PT (any grades and >= grade 3)
- Serious TEAEs
  - Serious TEAEs by SOC and PT
  - Serious TEAEs by PT
  - Serious treatment-related TEAEs by SOC and PT
  - Serious treatment-related TEAEs by PT
  - Serious tislelizumab -related TEAEs by SOC and PT
  - Serious ociperlimab/placebo -related TEAEs by SOC and PT
- TEAEs leading to death
  - TEAEs leading to death excluding death due to disease under by SOC and PT
  - TEAEs leading to death excluding death due to disease under by PT
  - TEAEs leading to death including death due to disease under by SOC and PT
  - TEAEs leading to death including death due to disease under by PT
  - Treatment-related TEAEs leading to death excluding death due to disease by SOC and PT
  - Treatment-related TEAEs leading to death excluding death due to disease by PT
- TEAEs leading to treatment discontinuation
  - TEAEs leading to treatment discontinuation by SOC and PT
  - TEAEs leading to treatment discontinuation by PT
  - TEAEs leading to tislelizumab discontinuation by SOC and PT
  - TEAEs leading to ociperlimab/placebo discontinuation by SOC and PT
  - Treatment-related TEAEs leading to treatment discontinuation by SOC and PT
  - Treatment-related TEAEs leading to treatment discontinuation by PT
- TEAEs leading to treatment modification
  - TEAEs leading to treatment modification by SOC and PT
  - TEAEs leading to treatment modification by PT
  - TEAEs leading to treatment modification of tislelizumab by SOC and PT
  - TEAEs leading to treatment modification of ociperlimab/placebo by SOC and PT
  - Treatment-related TEAEs leading to treatment modification by SOC and PT
  - Treatment-related TEAEs leading to treatment modification by PT

7.6.2.2 Immune-mediated Adverse Event

ImAEs are of special interest and summarized by category within a pre-defined list. The identification of imAEs will be described in immune-mediated adverse events charter. All imAE up to 90 days of last dose of any study drug, regardless of whether the patient starts a new anticancer therapy, will be summarized.

An overall summary table and separate summaries of the following incidence of imAEs will be provided:

- ImAEs by category and PT (any grades and >= grade 3)

- ImAEs leading to death by category and PT
- ImAEs leading to treatment discontinuation by category and PT
- ImAEs leading to treatment modification by category and PT
- ImAEs outcome, time to onset, and duration by category
- ImAEs treated with systemic corticosteroids by category

7.6.2.3 Infusion-related Adverse Event

For IRRs and IRRs with NCI-CTCAE grade ≥3, a summary of incidence by SOC and PT be provided.

7.6.2.4 Death

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

7.6.3. Laboratory Values

Laboratory safety tests will be evaluated for selected parameters described in Table 5, Abnormal laboratory values will be flagged and identified as those outside (above or below) the normal range. Reference (normal) ranges for laboratory parameters will be included in the CSR for this protocol.

Laboratory results will be summarized/listed using Système International (SI) units, as appropriate. Hematology and serum chemistry laboratory parameters that are graded in NCI-CTCAE v.5.0 will be summarized by shifts from baseline NCI-CTCAE grade to maximum post-baseline grades. In the summary of laboratory parameters by NCI-CTCAE grade, parameters with NCI-CTCAE grading in both high and low directions (e.g., calcium, glucose, magnesium, potassium, sodium) will be summarized separately. The summary tables will report lab assessments up to 30 days of the last dose date.

Laboratory parameters for potential Hy's Law for liver injury and abnormal thyroid function will also be summarized. Box-whisker plots will be generated for parameters of interest.

Patient data listings will be provided as appropriate.

Table 5: Clinical Laboratory Parameters

Clinical Chemistry	Hematology	Coagulation	Urinalysis	Thyroid Function
Alkaline phosphatase	Hematocrit	Prothrombin time	pH	TSH
ALT	Hemoglobin	Partial thromboplastin time or activated partial thromboplastin time	Specific gravity	Free T3

Clinical Chemistry	Hematology	Coagulation	Urinalysis	Thyroid Function
AST	Platelet counts	International normalized ratio	Glucose	Free T4
Albumin	White blood cell count		Protein	
Total bilirubin	Neutrophil count		Ketones	
Direct bilirubin	Lymphocyte count		Blood	
Blood urea nitrogen or urea			24-hour protein [a]	
Potassium				
Sodium				
Calcium				
Phosphorus				
Magnesium				
Chloride				
Creatinine				
Glucose				
Lactate dehydrogenase				
Total protein				
Creatine kinase/ CK-MB [b]				

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; CK-MB, creatine kinase-muscle/brain.

[a] On routine urinalysis, if urine protein is ≥ 2+ by dipstick then obtain a 24-hour urine sample for total protein or a random urine sample for total protein and creatinine to determine a protein-to-creatinine ratio.

[b] Cardiac enzyme testing has been added to monitor for potential event of immune-mediated myocarditis. If CK-MB fractionation is not available, assess troponin I and/or troponin T instead. Investigators should make every effort to perform either CK-MB, troponin I and/or troponin T consistently at screening and at follow-up visits.

7.6.4. Vital Signs

Vital signs will be listed by patient and visit.

7.6.5. Electrocardiograms (ECG)

The number and percentage of patients satisfying the following abnormal QTcF conditions at any time post-baseline will be summarized by treatment group:

- >450, > 480, or > 500 msec
- > 30 msec increase from baseline, or > 60 msec increase from baseline

7.6.6. Eastern Cooperative Oncology Group (ECOG) Performance Status

A shift table from baseline to worst post-baseline in ECOG performance score will be summarized.

7.7. Pharmacokinetic Analyses

Pharmacokinetic samples will be collected in this study as outlined in Appendix 1 of Protocol. Placebo samples will be collected but not analyzed.

The following analysis plan provides the framework for the summarization of the PK data from study AdvanTIG-203. The objective is to summarize available ociperlimab and tislelizumab PK concentrations following an IV administration. PK parameters will not be characterized as only sparse samples were collected.

Additional PK analyses, including population PK analyses and exposure-response analyses (efficacy or safety endpoints) may be conducted as appropriate and the results of such analysis may be reported separately from the CSR.

7.7.1. Reporting of Pharmacokinetic Concentrations for Descriptive Statistics

The ociperlimab and tislelizumab serum concentration data will be listed and tabulated by visit/cycle at which these concentrations are collected per the study design. Descriptive statistics will include means, medians, ranges, standard deviations, geometric means, and geometric CV%, as appropriate.

7.8. Immunogenicity Analyses

Anti-drug antibodies (ADAs) samples will be collected in this study as outlined in Appendix 1 of Protocol. Placebo samples will be collected but not analyzed.

The scope of ADAs calculations used for characterizing clinical immunogenicity depends on the incidence and kinetics of detected ADA. Therefore, not all parameters described below will be derived or additional parameters may be added. The effect of immunogenicity on PK, efficacy, and safety may be evaluated if data allows and will be reported separately from the CSR

The immunogenicity results will be summarized using descriptive statistics by the number and percentage of subjects who develop detectable ADAs for ociperlimab and tislelizumab separately based on Immunogenicity Analysis set. The incidence of positive and neutralizing ADAs (as applicable) will be reported for ADA-evaluable subjects according to the following definitions:

- **ADA-evaluable subject:** Number of subjects with reportable non-missing baseline result and at least one reportable sample taken after drug administration during the treatment or

follow-up observation period with reportable result (used for computing treatment induced ADA incidence).

- **Treatment-emergent ADA:** The sum of both treatment-boosted and treatment-induced ADA-positive subjects as a proportion of the evaluable subject population. Synonymous with "ADA Incidence".
- **Treatment-induced ADA:** ADA-evaluable subjects that were ADA-negative at baseline and ADA-positive following administration of biologic product.
- **Treatment-boosted ADA**: Baseline-positive ADA-evaluable subjects with significant increases (4-fold or higher) in ADA titer after biologic drug administration. Baseline-positive ADA-evaluable subject is an ADA-evaluable subject with positive ADA result at baseline.
- **Persistent ADA:** Treatment-induced ADA detected at two or more sampling time points during the 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 a period of 16 weeks or longer, or treatment induced ADA incidence only in the last sampling time point of the treatment study period or at a sampling time point with less than 16 weeks before an ADA-negative last sample.
- **Transient ADA:** Treatment-induced ADA that is not considered as persistent ADA.
- **Neutralizing ADA:** patients with positive NAb.

The individual immunogenicity results will also be listed.

8. INTERIM ANALYSES

No formal interim analyses will be conducted, and no formal stopping rules have been specified per protocol for this study. Summaries of efficacy and safety data may be generated to inform subsequent clinical development planning.

9. CHANGES IN THE PLANNED ANALYSISCHANGES 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

Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, Filiberti A, Flechtner H, Fleishman SB, de Haes JC, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst. 1993 Mar 3;85(5):365-76.

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

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

Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A on behalf of the EORTC Quality of Life Group. EORTC QLQ-C30 Scoring Manual (3rd edition). Brussels: EORTC, 2001

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

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.

Miettinen O, Nurminen M. Comparative analysis of two rates. Statistics in Medicine 1985; 4: 213–226.

Osoba D, Rodrigues G, Myles J, et al. Interpreting the significance of changes in health related quality-of-life scores. J Clin Oncol. 1998;16(1):139–144.

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

US Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER). FDA guidance for industry: clinical trial endpoints for the approval of cancer drugs and biologics. 2018. Accessed August 18, 2020.

Wen Y, Pan XF, Huang WZ, Zhao ZM, Wei WQ, Chen F, Lan H, Huang H, Yang CX, Qiao YL. Quality of Life for Patients with Esophageal/Gastric Cardia Precursor Lesions or Cancer: A One-year Prospective Study. Asian Pacific Journal of Cancer Prevention, Vol 16: 45-51, 2015.

11. APPENDIX

11.1. Appendix 1: Imputation Rules for Partial Dates

11.1.1. 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 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 > death date, then set to death date

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 > death date, then set to death 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.

11.1.2. 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, end of study date), then set to min (death date, end of study 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, the set to treatment start date
- If day is missing and month and year ≠ month and year of treatment start date, the 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.

11.1.3. Impute partial dates for subsequent anticancer surgery/procedure as collected in the post-treatment page

If start date of subsequent anticancer 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, start date of the subsequent anticancer therapy), then set to min (death date, study discontinuation date, data cutoff date, start date of the next subsequent therapy)
- The imputed start date must be before or equal to the end date.

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

11.1.4. Impute partial dates for 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
- 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.

11.2. Appendix 2: Rules For Identifying Missing Tumor Assessments

Identifying two missing tumor assessments

1) Input scheduled TA visit list
    a. (6wk-12wk-18wk-24wk-30wk-36wk-42wk-48wk-54wk-66wk-78wk…) for this study with TA as every 6 weeks for the first 54 weeks, then every 12 weeks thereafter
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. 6wk or 24wk) as –LPTADT_WK. It can be achieved programmatically by following the classification rule (e.g. defining thresholds) depicted in Table 6 below. (The team can consider to map all tumor visits if the scheduled visits code are uncleaned or questionable)
   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 (assuming 1 wk TA window) < earliest of PD/death date, then censor PFS at the –LPTADT
   b. Otherwise, PFS event at the earliest of PD/death date

Table 6 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 44 for an unscheduled visit, it will be mapped to Week 42 TA since it is within the Threshold for Week 42. Assuming it is SD and the subsequent TA of the patient is PD after Week 55, PFS will be censored at LPTADT (Week 44); had the PD occurred prior to Week 55, it would be counted as an PFS event.

Table 6: Example of Scheduled Tumor Assessments With Time Window

Weeks	Scheduled week -1	**Scheduled week**	Scheduled week+1	Threshold
Baseline		Baseline		
Every 6 weeks for the first 54 weeks	Week 5	Week6	Week 7	Week 9
	Week 11	Week 12	Week 13	Week 15
	Week 17	Week 18	Week 19	Week 21
	Week 23	Week 24	Week 25	Week 27
	Week 29	Week 30	Week 31	Week 33
	Week 35	Week 36	Week 37	Week 39
	Week 41	Week 42	Week 43	Week 45
	Week 47	Week 48	Week 49	Week 51
	Week 53	**Week 54**	Week 55	Week 60
Every 12 weeks thereafter	Week 65	**Week 66**	Week 67	Week 72
	Week 77	Week 78	Week 79	Week 84
	Week 89	Week 90	Week 91	…