STATISTICAL ANALYSIS PLAN

Study Protocol Number:	BGB-A317-311
Study Protocol Title:	A Phase 3, Randomized, Double-Blind, Placebo-Controlled Study of Tislelizumab (BGB-A317) Versus Placebo in Combination With Concurrent Chemoradiotherapy in Patients With Localized Esophageal Squamous Cell Carcinoma
Date:	Nov 14, 2024
Version:	2.0
NCT Number:	NCT03957590

SIGNATURE PAGE

Author [redacted]

[redacted]

Approval

[redacted]

[redacted]

[redacted]

TABLE OF CONTENTS

DOCUMENT REVISION HISTORY

Version	Date	Summary of Changes
1.0	Nov 18, 2021	
2.0	Nov 6, 2024	- Update document version and personnel information globally - Update Section 2, 6, 7 and 8 according to PA4.0 (30Sep2024) - Section 7.1: to add the rule for patients mistakenly take the other drug than the randomized for safety analysis set; update definition of PK/ADA analysis set according to PA4.0 - Section 7.2.4: update wordings of OS interim and final analyses according to PA4.0 - Section 7.3.2: remove analyses of critical protocol deviations due to updated standard process of protocol deviation - Section 7.3.3: remove the analyses of combined stratification factor - Section 7.3.4: add another category 'with non-target lesion only'; add other cutoffs of PD-L1 expression for exploration (e.g., 1%, 5%) - Section 7.3.8: remove analysis of post-treatment anti-cancer therapy duration - Section 7.4.1.2: remove the condition of Sensitivity Analysis 3 and will perform analysis regardless; add description of statistical test to assess proportional hazard assumption; modify prognostic factors to adjust in Supplementary Analysis 4, i.e., remove "histologic grade" and add "with non-target lesion only" - Section 7.4.1.3: remove the condition and will perform the analysis of discordance between IRC and investigator regardless - Section 7.4.2: reorganize analyses of OS as separate section and add sensitivity and supplementary analyses for final analysis - Section 7.4.3: remove analyses of QLQ-C30 index score - Section 7.4.4: remove "histologic grade", add "with non-target lesion only", add other cutoffs of PD-L1 expression for exploration (e.g., 1%, 5%) - Section 7.4.5 (7.4.4 in v1.0): removed as duplicated with Section 7.4.4 Subgroup Analysis - Section 7.5.1: clarify the first date of exposure as the first dose date of study treatment and add consideration of 'study discontinuation date' for last date of exposure to account for closeout analysis where data cutoff is not applicable - Section 7.5.2: TEAE added analyses of overview of treatment-related TEAE to any study drug, grade 3 or higher, serious, leading to death, TEAE leading to any treatment discontinuation and TEAE leading to

		any dose modification, as well as relevant summaries by SOC and PT; - Section 7.5.2: COVID-19 related remove analyses of AE related to COVID-19 by SMQ and PT, and added an overview of TEAE related to COVID-19; - Section 7.5.2: imAE adjust the time frame for imAE from "90 days after last tislelizumab/placebo" to "90 days after last study treatment"; adjust the analyses of imAE from "by category and preferred term (grade >=3 and all grades)" to "by category and worst grade"; adjust the analyses of imAE with grade 3 or higher from "by category, preferred term and worst grade" to "by category and worst grade"; combine previous analyses of outcome, time to onset, and duration together - Section 7.5.2: IRR reorganize the part to align with similar structure with TEAE and imAE - Section 7.5.4: remove summary of change from baseline for lab values - Section 7.5.5: remove summary of change from baseline for vital signs - Section 7.5.7: remove summary of ECOG by visit - Section 7.5.8: add analyses of AESI by treatment phase - Section 7.6: add more clarifications about the data to be included for PK analysis - Other minor adjustments and clarifications

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Term
ADA	Antidrug antibody
AE	Adverse event
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
AUC	Area under the curve
BGB-A317	Code name for monoclonal antibody tislelizumab
BIRC	Blinded Independent Review Committee
BOR	Best overall response
CBR	Clinical benefit rate
cCRT	Concurrent chemoradiotherapy
CI	Confidence interval
C_{max}	Maximum observed plasma concentration
C_{min}	Minimum observed plasma concentration
CR	Complete response
CSR	Clinical Study Report
CT	Computed tomography
CRT	Chemoradiotherapy
C_{trough}	Trough serum concentration
DCR	Disease control rate
DOR	Duration of response
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group

eCRF	Electronic case report form
EDC	Electronic data capture
EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30
EORTC QLQ-OES18	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Oesophageal Cancer Module.
ESCC	Esophageal squamous cell carcinoma
FDA	Food and Drug Administration
FOLFOX	5-fluorouracil, leucovorin, and oxaliplatin
GCP	Good Clinical Practice
HR	Hazard ratio
HRQoL	Health-related quality of life
ICF	Informed consent form
ICH	International Conference on Harmonisation
IDMC	Independent Data Monitoring Committee
imAE	Immune-mediated adverse event
IRT	Interactive Response Technology
ITT	Intent-to-Treat
IV	Intravenous(ly)
MedDRA	Medical Dictionary for Regulatory Activities
MinFU	Minimum study follow up
MRI	Magnetic resonance imaging
NA	Not applicable
NCI-CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
ND	No disease

NE	Not evaluated
ORR	Objective response rate
OS	Overall survival
PD	Progressive disease
PD-1	Programmed cell death protein-1
PD-L1	Programmed cell death protein ligand-1
PFS	Progression-free survival
PK	Pharmacokinetic(s)
PR	Partial response
PRO	Patient-reported Outcomes
Q3W	Once every 3 weeks
QTc	QT interval corrected for heart rate
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SD	Stable disease
SOC	System Organ Class
TEAE	Treatment-emergent adverse event
ULN	Upper limit of normal

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 BGB-A317-311: A Phase 3, Randomized, Double-Blind, Placebo-Controlled Study of Tislelizumab (BGB-A317) Versus Placebo in Combination With Concurrent Chemoradiotherapy in Patients With Localized Esophageal Squamous Cell Carcinoma. The focus of this SAP is for the planned primary, secondary, and exploratory analysis specified in the study protocol. This SAP is based on BGB-A317-311 protocol amendment Version 4.0 (PA v4.0) dated on 30 September 2024.

The analysis details for Pharmacodynamics, Pharmacogenomics and Biomarker analyses are not described within this SAP, and will be documented in separate analysis plans as appropriate.

PD-L1 expression is determined by PD-L1 score assessed by tumor area positive score (TAP), which is defined as the total percentage of the tumor area covered by tumor cells with any membrane staining above background and tumor-associated immune cells with any staining above background using Ventana PD-L1 (SP263) assay.

Reference materials for this statistical plan includes the protocol amendment v4.0 for BGB-A317-311 (dated as 30Sep2024) and Case Report Forms (Version 14.0,). If the protocol or case report forms are amended or updated, then appropriate adjustments to the SAP may be made if they are related to the planned analyses.

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

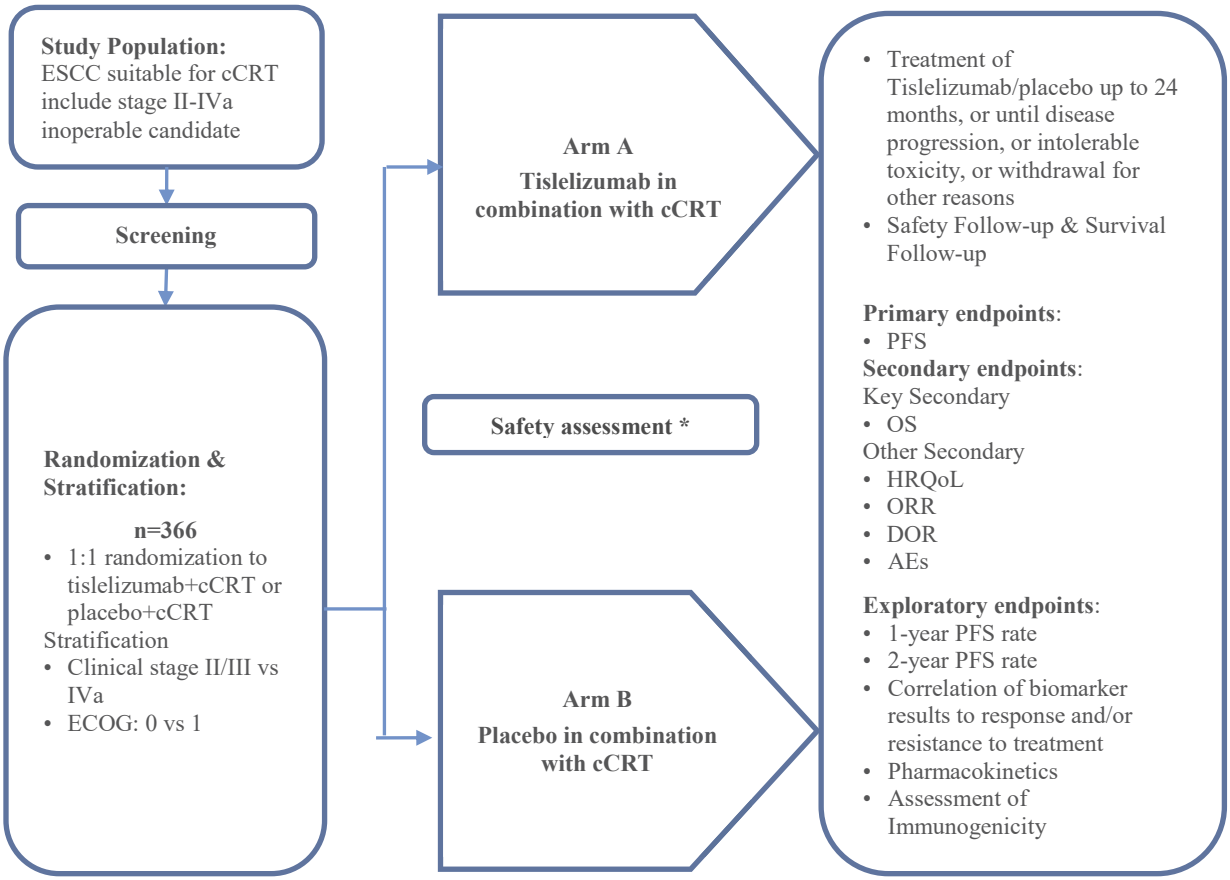
2 STUDY OVERVIEW

2.1 STUDY DESIGN

This is a double-blind, placebo-controlled, randomized, multicenter, Phase 3 study designed to evaluate the efficacy and safety of Tislelizumab versus placebo in combination with cCRT in patients with localized ESCC. This study is conducted in China. Patients with histologically confirmed localized ESCC who are considered suitable for cCRT (inoperable and without prior radiotherapy) are eligible.

The primary outcome measure of the study is progression-free survival (PFS) as assessed by the Blinded Independent Review Committee (BIRC) per Response Evaluation Criteria in Solid Tumors (RECIST) v1.1 and the key secondary outcome is overall survival (OS).

The study design schema is shown as below:



Abbreviations: AE, adverse events; cCRT, concurrent chemoradiotherapy; DOR, duration of response; ECOG, Eastern Cooperative Oncology Group; ESCC, esophageal squamous cell carcinoma; HRQoL, health-related quality of life; ORR, overall response rate; OS, overall survival; PFS, progression-free survival; vs, versus.
*Safety of Tislelizumab or placebo in combination with cCRT is assessed by an Independent Data Monitoring Committee after the first 20 patients (approximately 10 patients per each arm) have had at least 6 weeks of follow-up after the last dose of radiotherapy. Enrollment continues during the safety review.
For all study procedures, see Section 7 and Appendix 1 in PA v4.0.

Randomization of eligible patients will be stratified by the following two factors:

- Eastern Cooperative Oncology Group (ECOG) performance status: 0 versus 1
- Clinical stage: II/III versus IVa (AJCC version 8 [Rice et al 2017])

After 1:1 stratified randomization, patients will begin double-blind treatment with one of the following regimens:

- Arm A (study arm): Tislelizumab + chemotherapy + radiotherapy (concurrent)
- Arm B (control arm): Placebo + chemotherapy + radiotherapy (concurrent)

Crossover between treatment arms will not be allowed. And the treatment regimen could be detailed as below:

- Tislelizumab or placebo
    - 200 mg intravenously on Day 1 of every 21-day cycle (every 3 weeks) up to 24 months (about 35 cycles).

- Chemotherapy consists of Cisplatin and Paclitaxel
  - Cisplatin 25 mg/m^2 intravenously on Day 1 to 3 of every 21-day cycle (every 3 weeks) for 2 cycles.
  - Paclitaxel 135 mg/m^2 intravenously on Day 1 of every 21-day cycle (every 3 weeks) for 2 cycles.
- Radiotherapy
  - Radiotherapy will be delivered in both arms with the total dose of 50.4 Gy in 28 fractions and 5 fractions per week.

The safety of Tislelizumab or placebo in combination with cCRT is assessed by an Independent Data Monitoring Committee (IDMC) after the first 20 patients (approximately 10 patients per each arm) have had at least 6 weeks of follow-up after the last dose of radiotherapy and throughout the conduct of the trial. Enrollment continues during the safety review.

2.2 STUDY ASSESSMENTS

PFS and tumor response are assessed by BIRC and investigators. Baseline Tumor imaging (CT with or without contrast or magnetic resonance imaging [MRI]) and esophagography must be performed within 28 days prior to randomization. Tumor assessments after randomization will occur approximately every 9 weeks (± 7 days) for the first 54 weeks and then every 12 weeks (± 7 days) during Years 2 and 3, every 24 weeks (± 7 days) during Years 4 and 5, and then according to local standards with a minimum of 1 tumor response assessment annually thereafter, regardless of treatment delays. If a patient discontinues study treatment due to the reasons other than disease progression or death, tumor assessments will continue to be performed as scheduled until the start of subsequent anti-cancer therapy, experiences disease progression (confirmed by the investigator or BIRC, whichever is later), loss to follow up, withdrawal of consent, death, or until the study terminates, whichever occurs first. At the discretion of the investigator, patients may be treated beyond progression after obtaining agreement from the sponsor's medical monitor. After agreement is reached, the treating investigator must obtain written informed consent for treatment beyond radiologic disease progression and inform patients that this practice is not considered standard in the treatment of cancer.

Health-related quality of life (HRQoL) measured via patient-reported outcomes (PRO). PROs will be collected using the EORTC QLQ-C30 and EORTC QLQ-OES18 baseline, prior to dosing of every treatment cycle for the first 6 cycles, then every 2 cycles afterwards and at end of treatment. PROs will be completed prior to any clinical activities during on-study site visits.

After initiation of study treatment, all adverse events (AEs) and serious adverse events (SAEs), regardless of relationship to study treatment, will be reported until 30 days after last dose of study treatment or until the initiation of another anti-cancer therapy, whichever occurs first. Immune-mediated AEs (imAEs) will be recorded up to 90 days after the last dose of tislelizumab, regardless of whether the patient starts a new anti-cancer therapy. The investigator should report any SAEs that are believed to be related to tislelizumab treatment at any time after treatment discontinuation.

Safety and efficacy monitoring will be performed by IDMC. The IDMC may recommend modifications to the study, including study termination due to safety and/or efficacy concerns. The functions and membership of the IDMC will be described in an IDMC Charter.

3 STUDY OBJECTIVES

3.1 PRIMARY OBJECTIVES

- To compare the progression-free survival (PFS) as assessed by the Blinded Independent Review Committee (BIRC) per Response Evaluation Criteria in Solid Tumors (RECIST) v1.1 in the Intent-to-Treat (ITT) analysis set between tislelizumab in combination with concurrent chemoradiotherapy (cCRT) and placebo in combination with cCRT.

3.2 SECONDARY OBJECTIVES

Key Secondary Objective

- To compare the overall survival (OS) in the ITT analysis set between tislelizumab and placebo in combination with cCRT.

Other Secondary Objective

- To compare the patient reported outcomes of health-related quality of life (HRQoL) between tislelizumab and placebo in combination with cCRT.
- To compare the overall response rate (ORR) as assessed by the BIRC per RECIST v1.1 in the ITT analysis set between tislelizumab and placebo in combination with cCRT.
- To compare the duration of response (DOR) as assessed by the BIRC per RECIST v1.1 in the ITT analysis set between tislelizumab and placebo in combination with cCRT.
- To evaluate the safety and tolerability of tislelizumab in combination with cCRT.

3.3 EXPLORATORY OBJECTIVES

- To compare the 1-year PFS rate as assessed by the investigator per RECIST v1.1 in the ITT analysis set between tislelizumab and placebo in combination with cCRT.
- To compare the 2-year PFS rate as assessed by the investigator per RECIST v1.1 in the ITT analysis set between tislelizumab and placebo in combination with cCRT.
- To explore the potential predictive, prognostic biomarkers including but not limited to programmed cell death protein ligand-1 (PD-L1) expression; gene expression profiling (GEP); tumor infiltrating lymphocytes (TIL); tumor mutational burden (TMB), gene mutation, and microsatellite instability (MSI); and/or blood-based biomarkers (flow cytometry and immune-phenotyping [panel: A173, A167, A163, and A378], and circulating tumor DNA [ctDNA] alteration and peripheral blood mononuclear cell [PBMC]immune cell profiling) in archival and/or fresh tumor tissue and blood samples obtained before and/or during the study treatment and/or at disease progression, and the association with disease status, response to study treatment and mechanisms of resistance.
- To assess the pharmacokinetics (PK) of tislelizumab in combination with cCRT.
- To assess host immunogenicity to tislelizumab.

4 DEFINITION OF PRIMARY ESTIMAND AND KEY SECONDARY ESTIMAND

4.1 PRIMARY ESTIMAND

The primary clinical question of interest is: will tislelizumab in combination with concurrent

chemoradiotherapy treatment strategy prolong time to death/progression in locally advanced ESCC patients than concurrent chemoradiotherapy alone, had patients not been offered any new anti-cancer therapy that is not a part of assigned treatment strategy?

The justification for targeting this treatment effect is that we wish to estimate the relative effect of the two treatment strategies, in the absence of potentially confounding effect of any new anti-cancer therapy that is not a part of the assigned treatment strategy.

The primary estimand is described by the following attributes:

1. Treatment of interest:

   The **treatment of interest** is the randomized treatment (tislelizumab in combination with concurrent chemoradiotherapy or placebo in combination with concurrent chemoradiotherapy). Concurrent chemoradiotherapy includes Cisplatin, Paclitaxel and Radiotherapy.

2. Population:

   Adult patients with locally advanced inoperable ESCC, who are suitable for dCRT.

3. Primary variable:

   Progression free survival (see Section 5.1), defined as the time from randomization to the first documented disease progression, as determined by BIRC per RECIST v1.1, or death from any cause, whichever occurs first.

4. Handling of remaining intercurrent events:
   - Discontinuation of treatment: tumor assessment data collected after discontinuation of study treatment will be used for analysis (treatment policy strategy)
   - New anti-cancer therapy started prior to progression or death: data for patients who start to receive new anti-cancer therapy will be censored at the last tumor assessment date prior to the introduction of new therapy or randomization date if no valid tumor assessment before new anti-cancer therapy (hypothetical strategy)
   - Death due to COVID-19: death due to COVID-19 will be considered as part of PFS event (composite strategy)

5. Summary measure:

   Hazard ratio (HR) of PFS comparing tislelizumab in combination with concurrent chemoradiotherapy versus placebo in combination with concurrent chemoradiotherapy.

4.2 KEY SECONDARY ESTIMAND

The scientific question of interest is: will tislelizumab in combination with concurrent chemoradiotherapy treatment strategy prolong survival in locally advanced ESCC patients than concurrent chemoradiotherapy alone, regardless of whether subsequent anti-cancer therapy received.

The key secondary estimand is described by the following attributes:

1. Treatment of interest:

   The treatment of interest is the randomized treatment (tislelizumab in combination with concurrent chemoradiotherapy or placebo in combination with concurrent chemoradiotherapy)

2. Population:

   Adult patients with locally advanced inoperable ESCC, who are suitable for dCRT.

3. Variable:

   Overall survival (see Section 5.2), defined as the time from randomization to the date of death due to any cause.

4. Handling of remaining intercurrent events:
   - Discontinuation of treatment: any death or other data collected after discontinuation of study treatment will be used for analysis (treatment policy strategy)
   - New anticancer therapy started prior to death: any incidence will be ignored, ie. Any death or patients' data collected after the new anticancer therapy will be considered for analysis (treatment policy strategy).
   - Death due to COVID-19 infection will be counted as an event in the analysis of overall survival (composite strategy)
   - Any other unforeseen intercurrent events: all deaths and patients' data collected after any unforeseen intercurrent events will be considered for analysis (treatment policy strategy)
5. Summary measure:

   Hazard ratio (HR) of OS comparing tislelizumab in combination with concurrent chemoradiotherapy versus placebo in combination with concurrent chemoradiotherapy.

5 STUDY ENDPOINTS

5.1 PRIMARY ENDPOINTS

- PFS - defined as the time from randomization to the first documented disease progression, as determined by BIRC per RECIST v1.1, or death from any cause, whichever occurs first.

5.2 SECONDARY ENDPOINTS

Key Secondary Endpoints

- OS – defined as the time from the date of randomization to the date of death due to any cause.

Other Secondary Endpoints

- HRQoL assessment defined as patients' reported treatment effects using the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30 (EORTC QLQ-C30) and the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Oesophageal Cancer Module (EORTC QLQ-OES18).
- ORR – defined as the proportion of patients whose best overall response (BOR) is complete response (CR), or partial response (PR) as assessed by BIRC per RECIST v1.1.
- DOR – defined as the time from the first occurrence of a documented objective response to the time of relapse, as determined by the BIRC per RECIST v1.1, or death from any cause, whichever occurs first.
- The incidence and severity of treatment-emergent adverse events (TEAEs) graded according to National Cancer Institute-Common Terminology Criteria for Adverse Events (NCI-CTCAE, v5.0, 2017).

5.3 EXPLORATORY ENDPOINTS

- 1-year PFS rate – defined as the Kaplan-Meier estimator of proportion of patients alive and progression free at 1 year from randomization, as determined by investigator per RECIST v1.1.
- 2-year PFS rate – defined as the Kaplan-Meier estimator of proportion of patients alive and progression free at 2 years from randomization, as determined by investigator per RECIST v1.1.
- Status of PD-L1-related, immune-related, esophageal squamous cell carcinoma (ESCC)-related, and other exploratory biomarkers (GEP, TILs, TMB/gene mutation/MSI, and/or blood based biomarkers [flow cytometry and immune-phenotyping (panel: A173, A167, A163, and A378), ctDNA alteration, and PBMC immune cell profiling]) in archival and/or fresh tumor tissues and blood samples obtained before and/or during study treatment and/or at disease progression, and the association with disease status and/or response to tislelizumab in combination with cCRT and placebo in combination with cCRT. .
- Assessments of PK of tislelizumab when given with cCRT.
- Assessments of immunogenicity of tislelizumab by determining the incidence of anti-drug antibodies (ADAs).

6 SAMPLE SIZE CONSIDERATIONS

The sample size calculation is based on the number of events required to demonstrate the PFS superiority of Arm A over Arm B in the ITT analysis set. The estimates of the number of events required to demonstrate efficacy of PFS in the primary comparisons are based on the following assumptions:

- Median PFS of 14 months in Arm B.
- 1.5-month delayed treatment effect (i.e., assuming HR = 1 in the first 1.5 months; and HR = 0.65 thereafter, corresponding to an improvement in median PFS from 14 months to 21.5 months)
- One-sided α of 0.025
- 1-year drop-out rate of 5% for both treatment arms.
- Randomization ratio of 1:1.

With these assumptions, approximately 200 PFS events are required in the ITT analysis set for final analysis to obtain approximately 80% power. Approximately 366 patients will be enrolled, with enrollment duration of 24.0 months. The interim PFS analysis will be performed approximately 25.5 months after the first patient is randomized when 115 PFS events occurred, with error spent from a Lan-DeMets O'Brien-Fleming approximation spending function (See Section 8).

The key secondary endpoint OS will be tested only after the primary endpoint PFS is statistically significant. Based on the following assumptions: (1) median OS of 35 months in Arm B; (2) HR of 0.65, corresponding to an improvement in median OS from 35 months to 53.8 months, approximately 191 death events will provide approximately 85% power using 1-sided α of 0.025 to demonstrate the OS superiority of Arm A over Arm B in the ITT analysis set, with interim analysis for OS at the time of the interim PFS analysis (See Section 8).

The final analysis of PFS and OS will be performed concurrently when approximately 200 PFS events or approximately 191 OS events have been observed, whichever occurs first.

7 STATISTICAL METHODS

7.1 ANALYSIS SET

- The Intention-to-Treat (ITT) analysis set includes all randomized patients. It will be the primary analysis set for the efficacy analysis.
- The Safety analysis set includes all patients who received at least 1 dose of study treatment. It will be the primary analysis set for safety analysis. Patients in the safety analysis set will be classified according to treatment received, where treatment received is defined as (i) the intended treatment if it was received at least once, or (ii) the first treatment received if intended treatment is never received.
- The PK analysis set consists of all the patients who are randomized to the tislelizumab arm, and for whom postdose PK data are available. The PK analysis set will be used for PK analyses.
- The ADA analysis set includes all patients who are randomized to the tislelizumab arm and have a baseline and at least 1 postbaseline ADA result . The ADA analysis set will be used for ADA analyses.

7.2 DATA ANALYSIS GENERAL CONSIDERATIONS

Statistical programming and analyses will be performed using SAS® (SAS Institute, Inc., Cary, NC, USA), version 9.4 or higher, and/or other validated statistical software as required.

All descriptive statistics for continuous variables will be reported using mean, standard deviation (SD), median, 25 percentile (Q1), 75 percentile (Q3), minimum (Min), maximum (Max) and n. Categorical variables will be summarized as number (percentage) of patients.

The study Table, Listing, Figure shells will be in a separate document, which will show the content and format of all tables, listings, and figures in detail.

7.2.1 Definitions and Computations

Study day: For analysis of efficacy and baseline characteristics, study day will be calculated with reference to the date of randomization date, unless otherwise specified. For safety analysis, study day will be calculated with reference to the first dose date. For assessments conducted on or after the date of the first dose of study treatment/randomization date, study day will be calculated as (assessment date – date of first dose/randomization date + 1). For assessments conducted before the date of the first dose of study treatment/randomization date, study day is calculated as (assessment date – date of first dose/randomization date). There is no study day 0.

In the situation where the event dates are partial or missing, the dates will be presented as collected in the listings; Study day and any corresponding durations will be derived based on the imputed dates specified in Appendix 1.

Baseline Measurements:

- Baseline characteristics including ECOG and efficacy analysis except for HRQoL: Unless otherwise specified, a baseline value is defined as the last non-missing value collected prior to or at the time of randomization date.
- Safety analysis and lab assessments: Unless otherwise specified, a baseline value is defined as the last non-missing value collected prior to or at the time of first dose date.
- Vital signs and HRQoL: Unless otherwise specified, a baseline value is defined as the last non-missing value collected prior to or at the time of first tislelizumab or placebo dose date.

Study Follow-up Duration (SFD): Study follow-up duration is defined as the duration from the randomization date to the study discontinuation date (e.g. death, consent withdrawal, lost to follow-up) or to cutoff date for a patient who is still ongoing in the study.
Minimum study follow up (MinFU): Minimum study follow up is defined as a difference between the date of cut-off and the date of last patient randomized.

7.2.2 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 round to 0.0000 will be presented as '< 0.0001' and p-values that round to 1.000 will be presented as '> 0.9999'.
- Time-to-event or duration of event endpoints based on tumor assessment will be based on the actual date of the radiograph was obtained rather than the associated visit date.
- For lab results collected as < or >xxx, a numeric value, 0.0000000001 will be subtracted or added, respectively, to the value of xxx;
- 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).
- Retests and Unscheduled Visits: If not specified, unscheduled measurements will not be included in by-visit table summaries and figures but will contribute to best/worst case value where required (e.g. shift table). Listings will include scheduled, unscheduled and retest data.

7.2.3 Handling of Missing Data

Missing data will not be imputed unless otherwise specified elsewhere in the SAP. Missing dates or partially missing dates will be imputed conservatively for adverse events and prior/concomitant medications/procedures.

Specific rules for handling of missing or partially missing dates for adverse events and prior/concomitant medications/procedures are provided in Appendix 1.

Missing data for the HRQoL data will be handled according to each PRO instrument developer manual (Fayer & Machin, 2000 in The EORTC QLQ-C30 (Third Edition), 2001; https://euroqol.org/publications/user-guides/)

7.2.4 Multiplicity Adjustment

The family-wise error rate will be strongly controlled at 0.025 (1-sided) for the primary analysis of PFS as assessed by BIRC per RECIST v1.1 in the ITT analysis set.

There will be 1 interim analysis of PFS utilizing the Lan-DeMets O'Brien-Fleming approximation spending function. The stopping boundaries (P value) of the test for PFS at the interim and final analysis are shown in Section 8. The boundaries will be updated according to the actual numbers of events in the interim and final analyses.

To control overall type I error of the secondary endpoint testing, key secondary endpoint OS and other secondary endpoint HRQoL will be tested sequentially in ITT analysis set, the inferential test will be stopped at the first non-significant endpoint.

The key secondary endpoint of OS will only be tested after the primary endpoint is statistically significant. If it does not yield a significant result for OS interim analysis, the OS endpoint will be tested again at final analysis. A fixed 1-sided α level of 0.0001 will be allocated for the interim analysis of OS and the remaining 0.0249 will be allocated to final analysis as in Section 8.

HRQoL will only be tested after the key secondary endpoint OS is statistically significant. Bonferroni method will be applied to the test of OES-18 two symptoms dysphagia and eating using an alpha of 0.0125 for each symptom (0.025 totally as recycled from OS testing).

7.3 SUBJECT CHARACTERISTICS

7.3.1 Subject Disposition

The number and percentage of patients signed informed consent, randomization, screen failure, and screen previously will be summarized. The number and percentage of screen failure reason will also be summarized.

The number (percentage) of patients randomized, randomized but not treated, treated, discontinued from all study treatments, remained on any study treatment, remained on study, and discontinued from study will be summarized in the ITT analysis set.

For each study treatment, the number of patients randomized but untreated, treated, discontinued from treatments, remained on treatments will also be summarized in ITT analysis set. The number (percentage) of patients untreated with any concurrent chemoradiotherapy, treated with all concurrent chemoradiotherapy, and completed all chemoradiotherapy treatment will also be summarized.

The primary reasons for the last study treatment discontinuation, discontinuation of each study treatment, and study discontinuation will be summarized according to the categories in the CRF. Study follow-up time and minimum study follow-up time will be presented.

7.3.2 Protocol Deviations

Protocol deviation criteria will be established together with its category/term of important or non-important. Patients with important protocol deviations will be identified and documented before the database lock. Important protocol deviations will be summarized for all patients in the ITT analysis set.

Patient data listings of important and non-important protocol deviations will be provided.

Protocol deviations that are related to COVID-19 will be summarized.

7.3.3 Demographic and Other Baseline Characteristics

Demographic and other baseline characteristics will be summarized using descriptive statistics in the ITT analysis set.
Demographic and other baseline characteristics include:

- Age
- Age group (< 65 vs ≥ 65 years)

- Gender (male vs. female)
- Race
- Ethnicity
- BMI (kg/m^2)
- ECOG performance status at study entry
- Smoking status
- Alcohol consumption

In addition, the stratification factors per IRT and per CRF will be summarized:

- ECOG status (0 vs 1)
- Clinical stages (II/III vs IVa) (AJCC version 8 [Rice et al 2017])

7.3.4 Disease History and Baseline Disease Characteristics

The following disease history and baseline disease characteristics will be summarized in ITT population:

- Time from initial diagnosis to study entry (month)
- Prior chemotherapy
- Primary location
- Histologic grade
- Histologic type
- Clinical stage at study entry
- TNM stage at study entry
- Reason for inoperability
- With non-target lesion only
- PD-L1 expression status (PD-L1 score $\geq 10\%, < 10\ \%, \geq 5\%, < 5\%, \geq 1\%, < 1\%$, unknown)

Patient data listings of disease history and characteristics and cancer associated symptoms at baseline will be provided.

7.3.5 Prior Chemotherapy

The number (percentage) of patients with at least one prior chemotherapy, maximal cycle of chemotherapies (1,2,3), duration of prior chemotherapy, and time from end of last prior chemotherapy to study entry will be summarized by using ITT analysis set.

Prior chemotherapy will be coded using the World Health Organization Drug Dictionary (WHO drug) drug codes of the version currently in effect at Beigene at the time of database lock. Prior chemotherapy will be further coded to the appropriate Anatomical Therapeutic Chemical (ATC) code indicating therapeutic classification. The number (percentage) of patients reporting prior chemotherapy will be summarized by ATC medication class and WHO drug dictionary preferred term in the ITT analysis set.

7.3.6 Prior and Concomitant Medication

Prior and concomitant medications will be coded using the World Health Organization Drug Dictionary (WHO drug) drug codes of the version currently in effect at Beigene at the time of database lock. Prior and Concomitant medication will be further coded to the appropriate Anatomical Therapeutic Chemical (ATC) code indicating therapeutic classification.

The number (percentage) of patients reporting prior and concomitant medications, and reporting concomitant systemically administered corticosteroids/immunosuppressants will be summarized by ATC medication class and WHO drug dictionary preferred term in the safety analysis set. Prior medications are defined as medications that stopped before the first dose of study treatment. Concomitant medications are defined as medications that 1) started before the first dose of study treatment and were continuing at the time of first dose of study treatment, or 2) started on or after the date of the first dose of study treatment up to 30 days after the subject's last dose. In addition, systemic corticosteroid/immunosuppressant received between the first dose of study drug and up to 90 days after last dose will be summarized.
A listing of prior and concomitant medications will be provided.

7.3.7 Concomitant Procedure/Surgery

The number of patients receiving at least one concomitant procedure or surgery, type or name of procedure or surgery will be summarized. In case there is cancer-related surgery reported, cancer-related surgery (Yes or No), type or name of cancer-related surgery, and treatment intent of cancer-related surgery will also be summarized in safety analysis set.

7.3.8 Subsequent Anti-cancer Therapy

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

- As for efficacy analysis, start date of new anti-cancer therapy will be the earlist date of prohibited anti-cancer therapy taken during treatment, date of the post-treatment systemic anti-cancer therapy and 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 analysis is always the first date of new systemic anti-cancer therapy taken after the last study treatment.

ORR, PFS or OS benefit of Chinese herbal medicines or Chinese patent medicines (traditional Chinese medicine, TCM) has not yet been established. Therefore, TCM will not be considered as new anti-cancer therapy in the efficacy and safety analysis.

Subsequent anti-cancer therapy is defined as the anti-cancer therapy started after the last dose date of study treatment. A summary of number and pecentage of patients who received subsequent anti-cancer therapy in procedure or surgery/radiotherapy or systemic anti-cancer therapy/immunotherapy by treatment arm will be provided based on ITT analysis set.

The number (percentage) of patients by regimen number will be summarized. Time to first post-treatment anti-cancer therapy, and time to first post-treatment immunotherapy in systemic therapy and immunotherapy will be summaized descriptively. Patient data listings of post-treatment systemic therapy and post-treatment procedure/radiotherapy/surgery will be provided.

7.3.9 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 CRF, will be summarized by system organ class and preferred term in the ITT analysis set. A listing of medical history will be provided.

7.4 EFFICACY ANALYSIS

7.4.1 Primary Efficacy Endpoints

7.4.1.1 Primary analysis for primary estimand

Primary estimand is defined in Section 4.1.

Variable

Progression-Free survival is defined as the time from randomization to the first documented disease progression, as determined by BIRC per RECIST v1.1, or death from any cause, whichever occurs first. The PFS censoring rule will follow the FDA Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics (FDA, 2018).

It is worth noting that if the independent radiologists are not able to identify any disease at baseline (target or non-target lesions), they will be required to confirm that no lesions were identified at baseline. The overall tumor response assessment will be no disease at subsequent tumor assessment ((ND), provided the criteria for PD or NE are not met. In such case, ND or PD will be considered as valid tumor assessment when performing the analysis of PFS.

Detailed progression-free survival censoring rules for PFS primary and sensitivity analysis are described in Appendix 2.

PFS per the BIRC in ITT Analysis Set

The null and alternative hypotheses to be tested for PFS are:

$$H_0: PFS_T \leq PFS_C$$

$$H_a: PFS_T > PFS_C$$

where PFS_T and PFS_C represent the PFS of the treatment and the control groups, corresponding to tislelizumab in combination with concurrent chemoradiotherapy (cCRT) and placebo in combination with cCRT, respectively. The null hypothesis will be tested using a log-rank test stratified by the stratification factor at randomization as obtained via IRT (ECOG performance status: 0 versus 1, and clinical stage: II/III versus IVa). The stratified log-rank test will be the main estimator for the primary estimand of PFS. If the one-sided p-value is less than the pre-specified boundary, it will be concluded that the null hypothesis is rejected and the superiority of treatment group over the control group in PFS is demonstrated.

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 6 months, 12 months, 18 months, 24 months, 30 months and 36 months, etc. (if estimable), will be estimated using the Kaplan-Meier method for each treatment group. 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). 95% CIs for event-free rates will be estimated using Greenwood's formula (Greenwood, 1926).

Kaplan-Meier survival probabilities for each arm will be plotted over time. These analyses will be performed in the ITT analysis set.

7.4.1.2 Sensitivity analysis and supplementary analysis

Sensitivity analysis

Sensitivity Analysis 1: Unstratified PFS sensitivity analysis
The two treatment arms will be compared using the unstratified log-rank test. The HR together with the associated 95% confidence interval obtained using the unstratified Cox regression model will also be presented.

Sensitivity Analysis 2: To evaluate the impact of censoring rule of ≥ 2 consecutive missing or non-adequate tumor assessments on PFS
In the case where one or more tumor assessment is missing or non-adequate before progression/death, PFS events that occur after missing tumor assessments will be accounted. Otherwise the same analysis conventions as per the primary analysis will be used.

Sensitivity Analysis 3: PFS sensitivity analysis with stratification factors from the eCRF
PFS will be compared between the two treatment groups using the stratified log-rank test with stratification factors as collected in the eCRF. The HR together with the associated 95% confidence interval obtained using the stratified Cox regression model will also be presented, with stratification factors as collected in the eCRF.

Sensitivity Analysis 4: Investigator assessed PFS analysis
To support evidence on PFS assessed by BIRC, PFS by investigator will also be evaluated. The HR together with the associated 95% confidence interval obtained using the stratified Cox regression model will also be presented.

Sensitivity Analysis 5: PFS sensitivity analysis considering tumor response assessment no disease at baseline as not evaluable
To provide supportive evidence for primary analysis strategy, the overall tumor response assessment "no disease" (ND) will be considered similar as not evaluable (NE) or not applicable (NA), i.e., not adequate tumor assessment. Otherwise, the same analysis conventions as primary analysis.

Supplementary analysis

Supplementary Analysis 1: To evaluate the impact of new anti-cancer treatment on primary endpoint
This analysis targets an estimand which has the same attributes as the primary estimand except for handling the intervention effect. In this case, interest lies in the treatment effect regardless of whether new anticancer therapy received subsequently. Any new anticancer therapy will be ignored when derive PFS. Otherwise, the same analysis conventions as per the primary estimand will be used.

Analyses to assess proportional hazard assumption including Schoenfeld residual plot and time dependent covariate in the Cox model will be explored. While due to limitation of sample size, supplementary analyses 2 and 3 without non-proportional hazard assumption will be performed in case non-proportional hazard effect exists, regardless of the test result of proportional hazard assumption.

Supplementary Analysis 2: "PFS analysis based on Restricted Mean Survival Time method"
This analysis targets an estimand which has the same attributes as the primary estimand except the population level summary will be difference in RMST (Uno H, Claggett B, Tian L, Inoue E, et al. 2014)) between two treatment groups. In order to account for the possible non-proportional hazard effect, the RMST will be computed for PFS using the area under the curve from baseline to the minimum of the largest observed time on each of the two treatment groups. RMST will be computed for each treatment arm and the difference with its 95% CI will be displayed.

Supplementary Analysis 3: "PFS 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 combination of Fleming and Harrington weighted log-rank test (FH) test (Max-Combo, Satrajit R, Keaven A, Jiabu Y, Pralay M, 2019) based on the $G^{\rho,\gamma}$ family between two treatment groups (combination of $G^{0,0}$, $G^{0,1}$, $G^{1,1}$, $G^{1,0}$) to account for possibly different non-proportional hazard effects.

Supplementary Analysis 4: "PFS analysis adjusted for baseline covariates"
PFS will be also analyzed by adjusting multivariate covariates at baseline. The analysis addresses a different scientific question. i.e., will the addition of tislelizumab to concurrent chemoradiotherapy prolong PFS, adjusting for covariates. Covariate adjusted multivariate Cox regression model provides a conditional treatment effect rather than marginal treatment effect. A stratified Cox regression model will be performed with additional adjustment of key baseline prognostic factors as follows: prior chemotherapy (yes versus no), smoking status (never versus former/current), PD-L1 expression (PD-L1 score ≥ 10%, < 10%, unknown), age (< 65 versus ≥ 65), gender (female versus male), primary location, and with non-target lesion only (yes versus no), as appropriate.

7.4.1.3 Other supportive analysis

Discordance between IRC and investigator assessment

Discordance between IRC and investigator assessment on PD status or date will be summarized among patients read by IRC. Early discrepancy rate (EDR) and late discrepancy rate (LDR) will be calculated to detect potential evaluation bias of investigator assessment. More details could refer to Appendix 3.

7.4.2 Key Secondary Efficacy Endpoints

7.4.2.1 Primary analysis for key secondary estimand

Key Secondary estimand is defined in Section 4.2.

Variable

OS is defined as the time from the date of randomization to the date of death due to any cause. For patients who are alive by the clinical cutoff date, OS will be censored at the last known alive date (LKADT). 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 patients 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.

Overall Survival in ITT Analysis Set

The key secondary endpoint of OS will be tested only after the primary endpoint is statistically significant. The null and alternative hypotheses to be tested for OS are:

$$H_0 : OS_T \leq OS_C$$

$$H_a : OS_T > OS_C$$

where OS_T and OS_C represent the OS of the treatment and the control groups, respectively. The null hypothesis will be tested using a log-rank test stratified by the stratification factors at randomization (ECOG: 0 versus 1 and stage: II/III versus IVa). If the one-sided p-value is less than the pre-specified boundary, it will be concluded that the null hypothesis is rejected and the superiority of treatment group over the control group in OS is demonstrated.

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, will be estimated using the Kaplan-Meier method for each treatment group. 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). 95% CIs for event-free rates will be estimated using Greenwood's formula (Greenwood, 1926).

Kaplan-Meier survival probabilities for each arm will be plotted over time. These analyses will be performed in the ITT analysis set.

7.4.2.2 Sensitivity analysis and supplementary analysis

Sensitivity analysis

Sensitivity Analysis 1: Unstratified OS sensitivity analysis
The two treatment arms will be compared using the unstratified log-rank test. The HR together with the associated 95% confidence interval obtained using the unstratified Cox regression model will also be presented.

Sensitivity Analysis 2: OS sensitivity analysis with stratification factors from the eCRF
OS will be compared between the two treatment groups using the stratified log-rank test with stratification factors as collected in the eCRF. The HR together with the associated 95% confidence interval obtained using the stratified Cox regression model will also be presented, with stratification factors as collected in the eCRF

Supplementary analysis

Analyses to assess proportional hazard assumption including Schoenfeld residual plot and time dependent covariate in the Cox model will be explored. While due to limitation of sample size, supplementary analyses 2 and 3 without non-proportional hazard assumption will be performed in case non-proportional hazard effect exists, regardless of the test result of proportional hazard assumption.

Supplementary Analysis 1: “OS analysis based on Restricted Mean Survival Time method”
OS will be also analyzed by RMST method similar to PFS as mentioned in Section 7.4.1.2.

Supplementary Analysis 2: “OS analysis based on Max-Combo method”
OS will be also analyzed by Max-Combo method similar to PFS as mentioned in Section 7.4.1.2.

Supplementary Analysis 3: "OS analysis adjusted for baseline covariates"
OS will be also analyzed by adjusting multivariate covariates at baseline similar to PFS as mentioned in Section 7.4.1.2.

7.4.3 Other Secondary Efficacy Endpoints

Health-Related Quality of Life

The EORTC-QLQ-C30 consists of thirty questions that are specific to cancer and cancer treatment (Aaronson NK, et al., 1993; Fayers PM, et al., 2001). It includes one global health status/QoL (GHS) scale, five functional scales (physical, role, cognitive, emotional, and social), three symptom scales (fatigue, pain, and nausea and vomiting), and six single items (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 4 scales: dysphagia (3 items), eating (4 items), reflux (2 items), and pain (3 items) and 6 single item scales measuring saliva swallowing, choking, dry mouth, taste, coughing, and talking..

Higher scores in GHS and functional scales and lower scores in symptoms scales indicate better outcomes.

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.

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:

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

For symptom scales and global health status:

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

Refer to Table 1 and Table 2 for EORTC -QLQ-C30 and EORTC-QLQ-OES18 scoring.

OES18 index- score=∑ (DS of Dysphagia, DS of Eating, Reflux, Pain, Trouble swallowing saliva, Choked when swallowing, Dry mouth, Trouble with taste, Trouble with coughing, Trouble talking) ÷ # non-missing of DS

Table 1 OES18 Specific Symptoms Scales

	Scale	Number of items	Item range	OES18 Item Numbers
Symptom Scales				
Dysphagia	DY	3	3	1,2,3*
Eating	EA	4	3	6,7,8,9
Reflux	RE	2	3	14,15
Pain	PA	3	3	16,17,18
Single Items				
Trouble swallowing saliva	SA	1	3	4
Choked when swallowing	SW	1	3	5
Dry mouth	DM	1	3	10
Trouble with taste	TA	1	3	11
Trouble with coughing	CO	1	3	12
Trouble talking	TA	1	3	13

*: Reversing scoring items.

Table 2 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/ items				
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
Diarrhoea	DI	1	3	17
Financial Difficulties	FI	1	3	28

All HRQoL measures will be summarized in ITT analysis set.

Completion rates for the EORTC-QLQ-C30, and QLQ-OES18 will be summarized separately at each visit. A questionnaire module is considered complete and entered onto the analyses if at least one scale or one single item (1-item scale) is answered. In addition, the adjusted completion rates, defined as number of patients complete a PRO questionnaire divided by the number of patients still on treatment at the visit, will be summarized for each specific questionnaire.

For the EORTC-QLQ-C30, EORTC-QLQ-OES18 at each visit, raw score for functional scales and symptom scales will be calculated based on questionnaire items. Raw scores for the functional scale/symptom scale/single items will be transformed into 0-100 scale via linear transformation. The derived score (functional scales/symptom scales/single items and the global scale) of EORTC-QLQ-C30 and EORTC-QLQ-OES18, the index score of QLQ-OES18 will be summarized as well as change from baseline using descriptive analysis.

In addition, a mixed effect model analysis for measuring clinically meaningful changes from baseline will be performed using the key PRO endpoints of GHS, physical function, and fatigue domains of QLQ-C30 and dysphagia, reflux, pain and eating scales of QLQ-OES18. To assess treatment effects, change from baseline will be evaluated at the key clinical cycle 6 and cycle 10. The LS mean change from baseline (95% p values) for each arm and LS mean change difference (95% CI) between the arms from baseline, and the descriptive p values will be reported. Clinically meaningful difference is defined as ≈ ≥5 points difference from baseline and between the arms (Osoba 1998).

Bonferroni method of correction will be applied to the test of OES 18 symptoms of dysphagia and eating in clinically meaningful changes for cycle 6 using an alpha of 0.0125 for each symptom.

Time to clinically meaningful worsening will be analyzed for comparing the difference between the two treatment arms using Cox model for the key PRO endpoints (GHS, physical functioning, fatigue of the QLQ-C30 dysphagia, reflux, pain and eating, of QlQ-OES18), hazard ratio and its 95% confidence interval, and the descriptive p values will be provided, and a forest plot and KM curve for selected domains will be provided.

Time to clinically meaningful worsening is defined as the time from randomization to ≥ 10 points change from baseline in the worsening direction (Osoba 1998); i.e., -10 points from baseline in GHS and physical functioning scales, and +10 for symptom scales. A deterioration is not counted as an event if a subsequent improvement returned to baseline level or to less than 10 points. Patients without clinically meaningful worsening will be censored at the last PRO assessment.

KM estimates by treatment arm, the hazard ratio estimates, and their 95% CI will be provided for the key PRO endpoints. Hazard ratio is based on a Cox proportional hazard regression model, including treatment as covariate and stratified by ECOG status and clinical stages. P values will be reported, and except for dysphagia and eating that are formally tested, the p values for other PRO key endpoints will be descriptive.

Objective Response Rate by BIRC

Best overall response (BOR), defined as the best response recorded from randomization until progressive disease, or the start of new anticancer treatment. Patients with no post-baseline response assessment (due to any reason) will be considered non-responders for BOR.

ORR is defined as the proportion of patients whose BOR is complete response (CR) or partial response (PR) as assessed by BIRC per RECIST v1.1. Two-sided 95% CI of ORR will be calculated using Clopper-Pearson method in ITT analysis set. The odds ratio for ORR between treatment arms will be calculated using the Cochran-Mantel-Haenszel (CMH) method adjusting for ECOG performance status, and clinical stages, and its two-sided 95% CIs will be calculated. Mantel-Haenszel common risk difference in ORR will be estimated, with its 95% confidence interval constructed by a normal approximation and Sato's variance estimator. The proportion for each of the response categories (e.g. CR, PR, SD, PD, ND, NE, NA) will be presented by treatment arm.

Disease control rate (DCR) defined as the proportion of patients whose best overall response (BOR) is CR, PR, Non-CR/Non-PD or SD. DCR assessed by BIRC per RECIST v1.1 will also be summarized.

All the analyses will be based on confirmed results. Above analyses will also be repeated using assessments per investigator as sensitivity analysis.

Duration of Response by BIRC

Duration of response (DOR) is defined as the time from the first documented confirmed objective response to first documented disease progression or death, whichever occurred first. All the censoring rules for PFS primary analysis (Appendix 2**Error! Reference source not found.**) should be applied to DOR as well. DOR assessed by BIRC will be analyzed in the responders only.

The Median and other quantiles of DOR and the cumulative probability of DOR at 3, 6, 9, 12, 18, 24, and 30 months, etc. (if estimable), will be caculated using Kaplan-meier estimates for each treatment arm. The 95% confidence intervals for median and other quantiles will be estimated using a generalized Brookmeyer and Crowley method. And two-sided 95% confidence intervals for event-free rates will be estimated using Greenwood's formula.

7.4.4 Subgroup Analysis

Subgroup analysis of primary endpoint of PFS by BIRC and key secondary endpoint of OS will be conducted to assess the consistency of treatment effect across various subgroups. The unstratified HR estimates of PFS and OS, and the 95% CIs will be estimated within each category of the following variables: ECOG performance status (0 versus 1), age (< 65 versus >= 65 years), gender (female versus male), smoking status (never versus former/current), PD-L1 expression (PD-L1 score ≥ 10%, < 10 %, ≥ 5%, < 5%, ≥ 1%, < 1%, Unknown, as appropriate), primary location (cervical and upper thoracic versus middle thoracic versus lower thoracic and gastro-esophageal junction) , clinical stages (II/III versus IVa), prior chemotherapy (yes versus no), metastasis in regional lymph nodes (no versus yes), with non-target lesion only (yes versus no).

7.5 SAFETY ANALYSES

Safety will be assessed by monitoring and recording of all AEs graded by NCI-CTCAE v5.0. Laboratory values (e.g., hematology, clinical chemistry), vital signs, ECGs, and PEs, will also be used to evaluate safety. 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.

Safety data will be analyzed throughout treatment phase, unless otherwise selected analyses specified to be performed by concurrent phase and/or maintenance phase in Section 7.5.8.

7.5.1 Extent of Exposure

Extent of exposure will be summarized by treatment arm. And the extent of exposure to study treatment of tislelizumab/placebo, chemotherapy (by paclitaxel and cisplatin) and radiotherapy, will be summarized separately.

The following exposure parameters will be summarized with descriptive statistics for each study treatment. One cycle is defined as 21 days for tislelizumab/placebo and chemotherapy. Specifically:

Duration of exposure (month): the treatment duration will be calculated as (last date of exposure – first dose date + 1), and summarized in months for tislelizumab/placebo and days for chemotherapy and radiotherapy. For exposure of tislelizumab, duration of exposure will also be summarized in category of < 3 month, 3 - < 6 month, 6 - < 9 month, 9 - < 12 month, 12 - < 18 month, 18 - < 24 month, and ≥ 24 months.

- Duration of exposure will be calculated from the first dose date of study treatment.
- For study treatment of tislelizumab and chemotherapy, if patients discontinued treatment (with non-missing EOT date), "last date of exposure" is defined as min (cutoff date, study discontinuation date, death date, first dose date of the last cycle +

20). For radiotherapy, "last date of exposure" is defined as min (cutoff date, study discontinuation date, death date, last dose date).
- Otherwise for treatment ongoing patients, cutoff date is used as the "last date of exposure."

Number of cycles received: the number of cycles taken will be calculated as the sum of numbers of non-missing doses (dose>0) within each cycle for each study treatment for tislelizumab/placebo and chemotherapy. For number of cycles received in tislelizumab and placebo, sum of all cycles will be summarized, number of cycles received will also be summarized in category of 1-3, 4-6, 7-9, 10-12, 13-18, 19-24, 25-30, 31-36, and >36.

Number of fractions received: the number of fractions taken will be calculated as the sum of numbers of completed fractions in radiotherapy. Number of fractions received will also be summarized in category of 1-7, 8-14, 15-21, 22-27, 28 and >28.

Cumulative dose administered: the sum of all actual dose of study treatment, given from first to last administration. The summary will include drug administrations in all scheduled and unscheduled visits prior to the cutoff date. For study treatment of tislelizumab, unit of cumulative dose is mg. For study treatment of paclitaxel and cisplatin, unit of cumulative dose is mg/m^2. For study treatment of radiotherapy, unit of cumulative dose is Gy.

Actual dose intensity (ADI): Cumulative dose * 21 / (first dose date of the last cycle prior to or on cutoff date – first dose date + 21 days). For study treatment of tislelizumab, unit of actual dose intensity is mg/cycle. For study treatment of paclitaxel and cisplatin, unit of actual dose intensity is mg/m^2/cycle. See

Table 3 for calculation of ADI for paclitaxel and cisplatin. ADI will not be summarized for radiotherapy.

Relative Dose Intensity (%) = Actual Dose Intensity/Planned dose intensity *100%. Relative dose intensity will not be summarized for radiotherapy. For calculation reference,
Planned Dose Intensity for Tislelizumab (mg/cycle) = 200 mg/cycle
Planned Dose Intensity for Chemotherapy (mg/m^2/cycle): Cisplatin = 75 mg/m^2/cycle, Paclitaxel = 135 mg/m^2/cycle

Number of fractions received: the sum of all fractions received in radiotherapy. Number of fractions received will be summarized for radiotherapy only.
Radiotherapy Completion Rate: Radiotherapy completion rate is defined as the ratio of the cumulative dose (Gy) and the total planned dose of 50.4 (Gy).
Radiotherapy Completion Status: The summary of patients who completed radiotherapy, not completed radiotherapy but cumulative dose ≥ 90% planned doses, and not completed radiotherapy but cumulative dose < 90% of planned doses will be presented. 90% of planned doses is calculated based on total planned dose of 50.4 (Gy).

Delivery Over Planned Time: For radiotherapy, delivery over planned time was calculated as exposure duration of radiotherapy – exposure time of planned radiotherapy (date from the first fraction to the 28th fraction including fractions not done + 1).

For study treatment of tislelizumab, dose modification includes dose delay, infusion interrupted and infusion rate decreased. The number and percentage of patients with dose modification and the reason of dose modification due to AE and COVID-19 will be summarized.

For chemotherapy (paclitaxel and cisplatin), dose modification includes dose reduction, dose delay, infusion interrupted and infusion rate decreased. The number and percentage of patients with dose modification and the reason of dose modification due to AE and COVID-19 will be summarized.

For radiotherapy, dose modification includes dose delay. Number and percentage of patients with dose modification and the reason of dose modification due to AE, other reason in radiotherapy equipment related reason, administrative reason, and COVID-19 will be summarized.

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

Table 3 Example formulas to calculate actual dose intensity (ADI), planned dose and relative dose intensity (RDI) by cycle when target dose is in the unit of mg/m2

	ADI(mg/m^2/cycle)	Planned dose per cycle	RDI
cisplatin	$\frac{\sum_1^{\#\ of\ cycles} \frac{actual\ dose}{BSA*} \times 21}{date\ of\ first\ dose\ in\ last\ cycle + 21 - first\ dose\ date}$	25 X 3 mg/m^2	$\frac{ADI}{75}$
paclitaxel	$\frac{\sum_1^{\#\ of\ cycles} \frac{actual\ dose}{BSA*} \times 21}{date\ of\ first\ dose\ in\ last\ cycle + 21 - first\ dose\ date}$	135 mg/m^2	$\frac{ADI}{135}$

* to derive BSA at each visit is to use the formula present in Table 4.

Table 4 Example formulas to calculate Body Surface Area (BSA)

BSA(m^2)	Rounding
$BSA\ (m^2) = \sqrt{\frac{[height(cm) \times weight(kg)]}{3600}}$	None

* To derive BSA at each visit is to use baseline weight unless weight change for one visit is at least 10% greater compared to baseline weight. The general principle is to employ the same drug administration rule as the one specified in the Protocol.

7.5.2 Adverse Events

The AE verbatim descriptions (Investigator's description from the eCRF) will be coded using Medical Dictionary for Regulatory Activities (MedDRA). Adverse events will be coded to MedDRA (Version 27.0 or higher) by lower-level term (LLT) closest to the verbatim term. The linked MedDRA preferred term (PT), and primary system organ class (SOC) are also classified.

In this trial, a treatment-emergent adverse event (TEAE) is defined as an AE that had an onset date or a worsening in severity from baseline (pre-treatment) on or after the first dose of any study treatment up to 30 days after study treatment discontinuation, or initiation of new anti-cancer therapy, whichever occurs first. Only those AEs that were treatment-emergent will be included in summary tables.

All imAE will be reported separately as descriptions in Section 7.5.2.2.

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 per NCI-CTCAE v.5.0 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 TEAEs will also be summarized by relationship to the study treatment.

Treatment-related AEs include those events considered by the investigator to be related to study treatment or with missing assessment of the causal relationship.

All AEs, treatment emergent or otherwise, will be presented in listing.

7.5.2.1 Treatment Emergent Adverse Events

An overview summary of TEAEs will be summarized by the number (%) of patients with:

- At least one TEAE
    - At least one treatment-related TEAE
    - At least one treatment-related TEAE of tislelizumab/placebo
    - At least one chemotherapy related TEAE
    - At least one radiotherapy related TEAE
- At least one TEAE with grade 3 or higher
    - At least one treatment-related TEAE of grade 3 or higher
    - At least one treatment-related TEAE of tislelizumab/placebo of grade 3 or higher
    - At least one chemotherapy related TEAE of grade 3 or higher
    - At least one radiotherapy related TEAE of grade 3 or higher
- At least one serious TEAE
    - At least one treatment-related serious TEAE
    - At least one treatment-related serious TEAE of tislelizumab/placebo
    - At least one chemotherapy related serious TEAE

  - At least one radiotherapy related serious TEAE
- At least one TEAE leading to death *
  - At least one treatment-related TEAE leading to death *
  - At least one treatment-related TEAE of tislelizumab/placebo leading to death *
  - At least one chemotherapy related TEAE leading to death *
  - At least one radiotherapy related TEAE leading to death *
- At least one TEAE leading to any treatment discontinuation
  - At least one TEAE leading to tislelizumab/placebo treatment discontinuation
  - At least one TEAE leading to chemotherapy discontinuation
  - At least one TEAE leading to radiotherapy discontinuation
- At least one TEAE leading to any dose modification
  - At least one TEAE leading to tislelizumab/placebo dose modification
  - At least one TEAE leading to chemotherapy dose modification
  - At least one TEAE leading to radiotherapy dose modification

For patients with multiple occurrences of the same event will be counted only once, and the maximum grade per CTCAE v5.0 will be used. An overview of COVID-19 related adverse events will be summarized separately.

The incidence of following TEAEs will be reported by SOC and PT, sorted by decreasing frequency of SOC and PT in tislelizumab + CRT group:

- All TEAEs
- All TEAEs (≥3 and all grades)
- TEAEs with grade 3 or higher (≥3, grade 3, grade 4 and grade 5)
- Serious TEAEs
- TEAEs leading to death
  - TEAEs leading to death *
  - TEAEs leading to death including death due to disease under study
- TEAEs leading to treatment discontinuation
  - TEAEs leading to tislelizumab/placebo treatment discontinuation
  - TEAEs leading to chemotherapy treatment discontinuation
  - TEAEs leading to radiotherapy treatment discontinuation
- TEAEs leading to dose modification

  - TEAEs leading to tislelizumab/placebo dose modification
  - TEAEs leading to chemotherapy dose modification
  - TEAEs leading to radiotherapy dose modification
- Treatment related TEAEs
  - Treatment related TEAEs of any treatment
    - Treatment related TEAEs (≥3 and all grades)
    - Treatment related TEAEs with grade 3 or higher (≥3, grade 3, grade 4 and grade 5)
    - Treatment related serious TEAEs
    - Treatment related TEAEs leading to death *
    - Treatment related TEAE leading to death including death due to disease under study
  - Treatment related TEAEs of tislelizumab/placebo
    - Treatment related TEAEs of tislelizumab/placebo (≥3 and all grades)
    - Treatment related TEAEs of tislelizumab/placebo with grade 3 or higher (≥3, grade 3, grade 4 and grade 5)
    - Treatment related serious TEAEs of tislelizumab/placebo
    - Treatment related TEAEs of tislelizumab/placebo leading to death *
    - Treatment related TEAE of tislelizumab/placebo leading to death including death due to disease under study
  - Chemotherapy treatment related
    - Chemotherapy treatment related TEAEs (≥3 and all grades)
    - Chemotherapy treatment related TEAEs with grade 3 or higher (≥3, grade 3, grade 4 and grade 5)
    - Chemotherapy treatment related serious TEAEs
    - Chemotherapy treatment related TEAEs leading to death *
    - Chemotherapy treatment related TEAEs leading to death including death due to disease under study
  - Radiotherapy treatment related
    - Radiotherapy treatment related TEAEs (≥3 and all grades)
    - Radiotherapy treatment related TEAEs with grade 3 or higher (≥3, grade 3, grade 4 and grade 5)
    - Radiotherapy treatment related serious TEAEs

- Radiotherapy treatment related TEAEs leading to death *
- Radiotherapy treatment related TEAEs leading to death including death due to disease under study

* Only TEAE leading to death excluding death due to disease under study will be summarized.

Overall summary of all treatments exposure-adjusted event rate of TEAEs will be summarized. All Treatments exposure-adjusted event rate of all TEAEs will also be summarized.

In addition, TEAEs as follows will also be summarized by standardized MedDRA queries (SMQ) and PT:

1) Gastrointestinal Toxicity: Gastrointestinal toxicity is defined as any TEAE with narrow SMQ of gastrointestinal perforation, ulceration, hemorrhage, or obstruction.
2) Interstitial Lung Disease: Interstitial lung disease is defined as any TEAE with narrow SMQ of interstitial lung disease.
3) Any other events if necessary.

Patient data listings of all AEs will be provided.

7.5.2.2 Immune-mediated Adverse Events

Immune-mediated adverse events (serious or nonserious) were reported until 90 days after the last dose of study treatment regardless of initiation of a new anti-cancer therapy. Immune-mediated adverse events are of special interest and summarized by category within a pre-defined list as detailed in immune-mediated adverse events charter. The identification of immune-mediated adverse events is also described in immune-mediated adverse event charter..

Summaries of the following incidence of immune-mediated adverse events will be provided by treatment group:

- Overview of immune-mediated adverse events
- Immune-mediated adverse events by category, preferred term and worst grade
- Immune-mediated adverse events by category and worst grade
- Immune-mediated adverse events leading to death by category and preferred term
- Immune-mediated adverse events leading to treatment discontinuation by category and preferred term
- Immune-mediated adverse events leading to dose modification by category and preferred term
- Immune-mediated adverse events outcome, time to onset, and duration by category
- Immune-mediated adverse events treated with systemic corticosteroid by category
- Immune-mediated adverse events treated with hormone replacement therapy by category
- Immune-mediated adverse events treated with immunosuppressants by category

7.5.2.3 Infusion-related Adverse Event

Summaries of the following incidence of infusion related reaction will be provided by treatment group:

- Overview of infusion related reaction
- Infusion related reaction by SOC, PT and worst grade (≥3 and all grades)
- Infusion related reaction leading to tislelizumab/placebo treatment discontinuation by SOC and PT
- Infusion related reaction leading to dose modification of tislelizumab/placebo by SOC and PT

7.5.3 Death

All deaths and causes of death will be summarized by treatment group, including those occurred within 30 days and more than 30 days after last dose of study treatment.

Patient data listing of death and reason will be provided.

7.5.4 Laboratory Values

Clinical laboratory (e.g., hematology, serum chemistry, thyroid function, etc.) values and their changes from baseline will be evaluated for each laboratory parameter during the period defined for the treatment emergent adverse event as appropriate. Abnormal laboratory values will be flagged and identified as those outside (above or below) the normal range. Plots of laboratory values/change from baseline over time will be provided for selected lab parameters.

Laboratory parameters that are graded in NCI-CTCAE v.5.0 will be summarized by shift from baseline CTCAE grades 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., glucose, magnesium, potassium, sodium) will be summarized separately. Subjects with ≥ 2 grades increase in laboratory toxicities will also be summarized.

Hy's Law for liver injury and hepatic laboratory test will also be summarized. A listing of patients who meet Hy's law criteria will be provided.

7.5.5 Vital Signs

Plots of values and change from baseline over time will be provided for selected vital signs.

A listing of vital signs by patient and visit will be provided.

7.5.6 Electrocardiograms (ECG)

ECG will be performed at the baseline, Safety Follow-up, and as clinically indicated. The actual value and change from baseline of QTc intervals will be summarized by visit and treatment group using descriptive statistics as appropriate.

Abnormal post-baseline QTc 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.

A listing of ECG will be provided for all ECG recordings.

7.5.7 ECOG

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

A listing of ECOG performance status by visit will be provided.

7.5.8 Summary by Phase

Selected analyses will also be performed by concurrent phase and/or maintenance phase as supportive information.

7.5.8.1 Treatment phase

Summary of safety data in concurrent phase is based on safety analysis set; while summary of safety data in maintenance phase is based on the subset of safety analysis set, who received at least 1 dose of tislelizumab/placebo after date of last chemoradiotherapy administrated dose + 30 days.

Concurrent phase

1) If a patient has not received tislelizumab/placebo after the date of last chemoradiotherapy administrated dose + 30 days, the concurrent phase is defined from first dose date of any study treatment up to 30 days following study treatment discontinuation, or initiation of new anti-cancer therapy, whichever occurs first.
2) For the rest patients who have received at least 1 dose of tislelizumab/placebo after the date of last chemoradiotherapy administrated dose + 30 days, the concurrent phase is defined from first dose date of any study treatment up to the end date, which is (the first dose date of tislelizumab/placebo after (the date of last chemoradiotherapy administrated dose + 30 days)) – 1 day. Those patients will be considered ‘entered into maintenance phase’.

Maintenance phase

The maintenance phase is defined from the first dose date of tislelizumab/placebo after the date of last chemoradiotherapy administrated dose + 30 days, up to 30 days following the last study treatment discontinuation, or initiation of new anti-cancer therapy, whichever occurs first.

The analysis population and definition of concurrent/maintenance phase including start time and end time could be found in Table 5.

Table 5 Safety Summary by Phase

Phase	Patients	Phase Start Time	Phase End Time

Concurrent phase	1) If patient has not received tislelizumab/ placebo after the date of last chemoradiotherapy + 30 days	first dose of any study treatment	last dose of all study treatment + 30 days, or initiation of new anti-cancer therapy, whichever occurs first
	2) If a patient has received at least 1 dose of tislelizumab/placebo after the date of last chemoradiotherapy administrated dose + 30 days.		(the first dose date of tislelizumab/placebo after (the date of last chemoradiotherapy administrated dose + 30 days)) – 1 day
Maintenance phase	Patients entered into maintenance phase as defined above	first tislelizumab/ placebo administrated date after last chemoradiotherapy administrated dose + 30 days	30 days following the last study treatment discontinuation, or initiation of new anti-cancer therapy, whichever occurs first.

7.5.8.2 Selected analyses of TEAEs by treatment phase

TEAE in concurrent phase is defined as an AE that had an onset date or a worsening in severity from baseline (pre-treatment) in the time frame of concurrent phase. TEAE in maintenance phase is defined as an AE that had an onset date in the time frame of maintenance phase. The definition of concurrent/maintenance phase can be found in Table 5.

Selected types of TEAE as below will also be summarized by concurrent phase and/or maintenance phase:

- All TEAEs (≥3 and all grades)
- TEAEs with grade 3 or higher (≥3, grade 3, grade 4 and grade 5)
- Serious TEAEs
- TEAEs leading to death *
- TEAEs leading to death including death due to disease under study
- Gastrointestinal toxicity
- Interstitial lung disease

* Only TEAE leading to death excluding death due to disease under study will be summarized.

7.5.8.3 Selected analyses of imAEs by treatment phase

Selected types of imAE will also be summarized by concurrent phase and/or maintenance phase.

Immune-mediated AE in concurrent phase is defined as an imAE that had an onset date or a

worsening in severity in the concurrent phase defined by Table 5, with phase end time extending to 90 days after following the last study treatment discontinuation, regardless of new anti-cancer therapy, for patients have not received tislelizumab/placebo after the dose of last chemoradiotherapy administration dose + 30 days; or as Table 5 defined for patients have received at least 1 tislelizumab/placebo after the dose of last chemoradiotherapy administration dose + 30 days.

Immune-mediated AE in maintenance phase is defined as an imAE that had an onset date in the maintenance phase defined by Table 5 with phase end time extending to 90 days following the last study treatment discontinuation, regardless of new anti-cancer therapy.

Selected types of imAE as below will also be summarized by concurrent phase and/or maintenance phase:

- Immune-mediated adverse events by category, preferred term and worst grade (≥3 and all grades)
- Immune-mediated adverse events by category and worst grade
- Immune-mediated adverse events leading to death by category and preferred term

7.5.8.4 Selected analyses of IRRs by treatment phase

For infusion related reaction (IRR)s, a summary of incidence by SOC, PT and maximum severity by treatment phase will also be provided.

7.5.8.5 Other safety analyses by treatment phase

For selected safety value as below, safety value reported in time frames according to Table 5 will be summarized by treatment phase:

- Summary of hematology / serum chemistry: Shifts from baseline to the worst post-baseline grade
- Summary of hematology / serum chemistry: Increase in 2 or more CTCAE toxicity grades as compared with baseline
- Summary of hepatic laboratory test
- Summary of laboratory tests for potential Hy's law
- Summary of thyroid laboratory test
- Summary of ECOG performance status: Shifts from baseline to the worst post-baseline grade
- Summary of ECG
- Abnormal postbaseline QTc results

7.6 PHARMACOKINETIC ANALYSES

Pharmacokinetic samples will be collected in this study as outlined in Protocol Appendix 1, and only from patients randomized to receive BGB-317 in sites that are able to adequately perform PK sampling, handling, and processing procedures as outlined in the Laboratory Manual.

Tislelizumab postdose and trough serum concentration (C_{trough}) data will be tabulated and summarized by visit/cycle at which these concentrations are collected. Descriptive statistics will include means, medians, ranges, and standard deviations, as appropriate.

All subjects in PK analysis set will be included for data analysis. For subjects in the tislelizumab arm take the placebo drug, if the following PK samples are collected within 12 weeks, these PK results will be removed from the descriptive statistics.

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

7.7 IMMUNOGENICITY ANALYSES

Human anti-drug antibodies (ADA) to tislelizumab will be assessed during the study as defined in the protocol.

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 negative at baseline and ADA positive post-baseline.
  - Persistent ADA response is defined as Treatment-induced ADA detected at 2 or more time points during treatment or follow-up, where the first and last ADA positive samples are separated by 16 weeks or longer; or detected in the last time point.
  - Transient ADA response is defined as a treatment-induced response that is not considered persistent.
- Neutralizing ADA is defined as ADA that inhibits or reduces the pharmacological activity.

ADA response endpoints:

- ADA incidence is defined as sum of treatment-emergent ADA, which include both treatment-induced and treatment-boosted ADA-positive patients, as a proportion of the ADA evaluable population.
- ADA prevalence is defined as 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 ADA. The incidence of positive ADA and

neutralizing ADA will be reported for evaluable patients. The effect of immunogenicity on PK, efficacy, and safety may be evaluated if data allowed.

8 INTERIM ANALYSIS

PFS

One interim analysis is planned for PFS when 115 targeted PFS events are observed.

The efficacy boundary is estimated based on the Lan-DeMets O'Brien-Fleming approximation spending function. The stopping boundaries (P value) of the test for PFS at the interim and final analysis are shown in Table 6. The boundaries will be updated according to the actual numbers of events in the interim and final analyses, to ensure overall type I error controlled.

Table 6: The Analysis Timing and Efficacy Boundary for PFS in the Interim Analysis and Final Analysis

PFS	Estimated Timing (months)	Estimated # of Events Observed	P-value [a] for Efficacy	Approximate HR threshold
Interim Analysis	25.5	115	< 0.0062	< 0.627
Final Analysis [b]	45.1	200	< 0.0224	< 0.753

Abbreviations: HR, hazard ratio
[a] one-sided
[b] If target number of PFS events are not observed when there have been approximately 191 OS events, the PFS final analysis will be performed at that time, the PFS final analysis will be performed at the time, regardless of the number of PFS events observed.

OS

The key secondary endpoint of OS will be tested only after the primary endpoint is statistically significant. If the test of PFS is significant, the total α=0.025 will be allocated to OS interim and final analyses to test for significance.

The interim analysis of OS was planned at the time of the PFS interim analysis, with a fixed 1-sided α level of 0.0001 allocated, and the final analysis of OS will be performed at the same time of the PFS final analysis, with remaining fixed 1-sided α of 0.0249.

The detailed testing strategy which controls overall type I error rate at 2.5% is described as follows:

1) If the interim analysis for PFS is significant, OS interim analysis will be performed with 1-sided α of 0.0001. If OS is not significant at the OS interim analysis, final analysis for OS will be performed at the same time of PFS final analysis with 1-sided α of 0.0249.

2) If the PFS IA is not significant, PFS will continue to final analysis. If the final analysis for PFS is positive, an alpha of 0.025 will be propagated to OS, which will be allocated as aforementioned and OS final analysis will be performed with 1-sided α of 0.0249.

3) If PFS continues to final analysis and is not significant, OS will not be statistically evaluated.

An IDMC might make the recommendation regarding stopping the study early for compelling efficacy results. More details will be given in the IDMC 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

Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, Filiberti A, Flechtner H, Fleishman SB, de Haes JCJM, Kaasa S, Klee MC, Osoba D, Razavi D, Rofe PB, Schraub S, Sneeuw KCA, Sullivan M, Takeda F. The European Organisation for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. Journal of the National Cancer Institute 1993; 85: 365-376.

Bergman B, Aaronson NK, Ahmedzai S, Kaasa S, Sullivan M (1994) The EORTC QLQ-LC13: a modular supplement to the EORTC Core Quality of Life Questionnaire (QLQ-C30) for use in lung cancer clinical trials. EORTC Study Group on Quality of Life. Eur J Cancer 30A: 635–642.

Clopper, C. and Pearson, ES. The use of confidence or fiducial limits illustrated in the case of the binomial, Biometrika 1934; 26: 404-413.

Common Terminology Criteria for Adverse Events (CTCAE) Version 5.0. US Department of Health and Human Services.
https://ctep.cancer.gov/protocoldevelopment/electronic_applications/docs/CTCAE_v5_Quick_Reference_8.5x11.pdf. Published 27 November 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. Brussels: EORTC, EORTC QLQ-C30 Scoring Manual (3rd

edition). 2001. ISBN: 2-9300 64-22-6.

Greenwood M. “The natural duration of cancer”. Reports on Public Health and Medical Patients (London: Her Majesty’s Stationery Office). 1926; 33:1-26.Groenvold M, Klee MC, Sprangers MA, Aaronson NK (1997) Validation of the EORTC QLQ-C30 quality of life questionnaire through combined qualitative and quantitative assessment of patient-observer agreement. J Clin Epidemiol 50: 441–450.

Kaplan E, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc. 1958;53:457-81.

Mantel, Nathan, and William Haenszel. 1959. “Statistical Aspects of the Analysis of Data From Retrospective Studies of Disease” 22: 30.Maringwa J, Quinten C, King M, et al. Minimal clinically meaningful differences for the EORTC QLQ-C30 and EORTC QLQ-BN20 scales in brain cancer patients. Annals of Oncology 2011; 22: 2107–2112.Osoba D, Zee B, Pater J, Warr D, Kaizer L, Latreille J (1994) Psychometric properties and responsiveness of the EORTC quality of Life Questionnaire (QLQ-C30) in patients with breast, ovarian and lung cancer. Qual Life Res 3: 353–364.
Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. Stat Med. 1998;17:2815-34.

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.

Rice TW, Ishwaran H, Ferguson MK, et al. Cancer of the esophagus and esophagogastric junction: an eighth edition staging primer. J Thorac Oncol. 2017;12(1):36-42.

Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol. 1998 Jan;16(1):139-44. doi: 10.1200/JCO.1998.16.1.139. PMID: 9440735.

Sato, Tosiya, Sander Greenland, and James M. Robins. 1989. "On the Variance Estimator for the Mantel-Haenszel Risk Difference." *Biometrics* 45 (4): 1323–24.

Uno H, Claggett B, Tian L, Inoue E, Gallo P, Miyata, et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis J Clin Oncology. 2014; 32(22):2380-5.US Food and Drug Administration Center for Drug Evaluation Research (CDER) and Center for Biologics Evaluation and Research. FDA Guidance for Industry: Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. 2018.

APPENDIX

APPENDIX 1 MISSING DATA IMPUTATION

Handling of Missing/Partially Missing Dates

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

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 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.

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, 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, 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

- If the imputed end date > min (death date, study discontinuation date), then set to min (death date, study discontinuation date)

1.3 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 THE PRIMARY AND SELECTED SUPPLEMENTARY/SENSITIVITY CENSORING RULES FOR THE DERIVATION OF PFS

No.	Situation	Date of Event or Censoring	Primary Analysis	Supplementary Analysis 1	Sensitivity Analysis 2
1	No baseline or any post-baseline tumor assessments and without death within 19 weeks from reference start date	Reference start date	Censored	Censored	Censored
2	No baseline or any post-baseline tumor assessments and died within 19 weeks from reference start date	Date of death	Event	Event	Event***
3	Progression documented between scheduled visits	Date of first radiologic PD assessment	Event	Event	Event***
4	No progression or death at the time of data cut-off or withdrawal from study or lost to follow up or other EOS reasons	Date of last adequate radiologic assessment prior to or on date of data cut-off or withdrawal from study or lost to follow up or other EOS reasons*	Censored	Censored	Censored
5	New anticancer treatment started	Date of last adequate radiologic assessment prior to or on date of new anticancer treatment*	Censored	Event (on PD or death date)	Censored
6	Death before first PD assessment	Date of death	Event	Event	Event***
7	Death or progression after two or more consecutive missed visits**	Date of last adequate radiologic assessment before missed tumor assessments*	Censored	Censored	Event (on PD or death date)***

*Adequate tumor assessment is a radiologic assessment of CR, PR, SD, non-CR/non-PD, ND (only applicable for BIRC) or PD as determined by the reviewers. Clinical PD without documented radiologic assessment is not considered as PD event, unless otherwise specified by protocol.

** Two (or more) consecutive missing visits are identified as described in Section “Identifying two (or more) consecutive missing tumor assessments” below.

*** Date for PFS event will be the earliest date of events defined in 2,3,5,6,7.

The reference start date is the randomization date.

The priority of the censoring rules in the primary analysis is as follows:

1. If the patient had PD or death, the following sequence will be applied:

a. If a patient did not have either baseline or any post-baseline tumor assessment (No. 1), the patient will be censored on the reference start date. However, if the patient died within the first 2 scheduled assessment time windows after reference start date and did not receive new anticancer treatment, then the patient will be counted as event and the date of death will be the PFS event date (not censored) (No. 2).

b. If a patient had new anticancer treatment before PD or death (No. 5), the patient will be censored on the date of the last adequate tumor assessment prior to or on the date of new anticancer treatment.

c. If a patient missed two or more consecutive assessments before PD or death (No. 7), the patient will be censored on the date of the last tumor assessment before PD or death. Note that if a patient is censored by both this criterion and the anticancer treatment criteria, the earliest censoring date will be used and the censoring reason will be receiving new anticancer treatment.

d. Otherwise, if a patient had an event (No. 3, No. 6.), the earliest event date will be used.

2. If a patient did not have PD or death, the censoring date will be the earliest censoring date if the patient met multiple censoring criteria (No. 1, No. 4, No. 5, No. 7).

3. In sensitivity analysis 1, the PFS event date will be derived ignoring new anti-cancer therapy (i.e., this intercurrent event is handled using treatment policy strategy).

4. In sensitivity analysis 2, Any PD or death after two or more consecutive missing tumor assessment will be considered as a PFS event (i.e., this intercurrent event is handled using treatment policy strategy).

Identifying two (or more) consecutive missing tumor assessments:

1) Input scheduled TA visit list for the study as shown in below table
2) Identify last evaluable TA before PD or death (LPTADT) and map it to the closest scheduled visit (LPTADT_WK).
3) In the event of unscheduled TA, choose the closest scheduled visit number (e.g. 27wk) as LPTADT_WK. It can be achieved programmatically by following the classification rule (e.g. defining thresholds) depicted in table below. It can be considered to map all tumor visits if the scheduled visits code are uncleaned or questionable. Otherwise, assign the scheduled visit number (assuming it is coded correctly) to LPTADT_WK
4) Find the 2nd TA visit after LPTADT_WK according to the scheduled TA list in step 1) (LPTADT_WK_2)
    a. If LPTADT_WK_2+1wk (assuming 1 wk TA window) < earliest PD/death date, then there are two or more consecutive missing TAs and will censor PFS at LPTADT

Otherwise, it will be counted as PFS event at the earliest of PD/death date.

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 2 above). For example, if LPTADT is Week 42 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 (ie, 2 consecutive missing assessments), PFS will be censored at LPTADT (Week 42); had the PD occurred prior to/at Week 67, it would be counted as an PFS event.

Weeks	Scheduled week -1	**Scheduled week**	Scheduled week+1	Threshold
Baseline		Baseline		
Every 9 weeks for the first 54 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
	Week 53	Week 54	Week 55	Week 60
Every 12 weeks during Years 2 and 3 (~156 weeks)	Week 65	Week 66	**Week 67**	Week 72
	Week 77	Week 78	Week 79	Week 84
	Week 89	Week 90	Week 91	Week 96
	Week 101	Week 102	Week 103	Week 108
	Week 113	Week 114	Week 115	Week 120
	Week 125	Week 126	Week 127	Week 132
	Week 137	Week 138	Week 139	Week 144
	Week 149	Week 150	Week 151	Week 162
Every 24 weeks during Years 4 and 5 (~260 weeks)	Week 173	Week 174	Week 175	Week 186
	Week 197	Week 198	Week 199	Week 210
	Week 221	Week 222	Week 223	Week 234
	Week 245	Week 246	Week 247	Week 270
Every 48 weeks afterwards	Week 293	Week 294	Week 295	Week 318
	Week 341	Week 342	Week 343	Week 366
		…		

Note: the “Threshold” column is just for mapping the last evaluable TA before PD/death, while the “scheduled week+1” column should be used as the upper limit boundary for the 2 consecutive missing (scheduled) visits.

APPENDIX 3 THE EARLY DISCREPANCY RATE (EDR) AND LATE DISCREPANCY RATE (LDR) CALCULATION METHOD

PFS concordance between investigator assessed PD and BIRC assessed PD will be summarized as below,

		BIRC	
		PD	No PD
Investigator	PD	a = a1 + a2 + a3	b
	No PD	c	d

a1: number of agreements on timing and concurrence of PD

a2: number of times investigator declares PD later than BIRC

a3: number of times investigator declares PD earlier than BIRC

The early discrepancy rate (EDR) is defined as proportion of investigator declares PD earlier than IRC among all investigator assessed PDs, i.e.,

$$EDR = \frac{b + a3}{a + b}$$

Th late discrepancy rate (LDR) is defined as proportion of investigator declares PD later than IRC among all discrepant cases including asynchronous PDs, i.e.,

$$LDR = \frac{c + a2}{b + c + a2 + a3}$$

The differential discordance of EDR and LDR between two arms will be calculated as the rate of tislelizumab+CRT arm minus the rate of placebo+CRT arm.