BeiGene

STATISTICAL ANALYSIS PLAN

Study Protocol Number:	BGB-A317-302
Study Protocol Title:	A Randomized, Controlled, Open-label, Global Phase 3 Study Comparing the Efficacy of the anti-PD-1 Antibody Tislelizumab (BGB-A317) versus Chemotherapy as Second Line Treatment in Patients with Advanced Unresectable/Metastatic Esophageal Squamous Cell Carcinoma
Date:	Dec 22, 2020
Version:	3.0

SIGNATURE PAGE

Author: Aiyang Tao, Ph.D

Director, Statistics and Data Science

DocuSigned by:
Aiyang Tao
Signer Name: Aiyang Tao
Signing Reason: I approve this document
Signing Time: 27-Dec-2020 | 11:52:18 PST
05A28BACD928482DAAC9B814B4D311E0

Signature

Date: _____

Approval

Xiang Guo, Ph.D.

Vice President, Statistics and Data Science

DocuSigned by:
xiang guo
Signer Name: xiang guo
Signing Reason: I approve this document
Signing Time: 25-Dec-2020 | 07:59:08 PST
C4979E19F5584C92B3FD70F23DE589AA

Signature

Date: _____

Liyun Li, MD

Associate Director, Clinical Development

DocuSigned by:
liyun li
Signer Name: liyun li
Signing Reason: I approve this document
Signing Time: 24-Dec-2020 | 06:44:12 PST
12A42D5D75584360BAD27AF981A0B194

Signature

Date: _____

Vivian Huang, MD

Sponsor Medical Monitor

DocuSigned by:
wei huang
Signer Name: wei huang
Signing Reason: I approve this document
Signing Time: 27-Dec-2020 | 16:54:07 PST
CA7D9EA2F8404C448FB31048385C588C

Signature

Date: _____

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Term
ADA	Antidrug antibody
AE	Adverse event
BGB-A317	Tislelizumab, a humanized monoclonal antibody directed at PD-1
BOR	Best overall response
C_{max}	Maximum observed plasma concentration
CMH	Cochran-Mantel-Haenszel
CR	Complete response
CT	Computed tomography
C_{trough}	Trough serum concentration
DCR	Disease control rate
DOR	Duration of response
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
CRF	Case report form
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-Oesophagus Cancer Module
EQ-5D-5L	European Quality of Life 5-Dimensions (EQ-5D-5L Version)
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HRQoL	Health Related Quality of Life
ICC	Investigator chosen chemotherapy
imAE	Immune-related adverse event
IRT	Interactive response technology
ITT	Intention-to-Treat
KM	Kaplan-Meier
MedDRA	Medical Dictionary for Regulatory Activities

MSI	Microsatellite instability
NCI-CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
ORR	Overall 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)
PP	Per-Protocol Population
PR	Partial response
PT	Preferred term
Q2W	Once every two weeks
Q3W	Once every three weeks
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
US	United States

1 INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to describe the procedures and the statistical methods that will be used in data analysis and results reporting for A317-302: A randomized, controlled, open-label, global Phase 3 study comparing the efficacy of the anti-PD-1 antibody tislelizumab (BGB-A317) versus chemotherapy as second line treatment in patients with advanced unresectable/metastatic esophageal squamous cell carcinoma. This SAP is based on BGB-A317-302 Protocol Amendment 4.0 dated as March 20th, 2020. The focus of this SAP is the planned final analysis specified in the study protocol. The study was masked to statistician and the statistical analysis plan is finalized before unmasking.

The analysis details for pharmacokinetic (PK), pharmacodynamics, pharmacogenomics and biomarker analyses are not described within this SAP. Separate analysis plans will be completed for these analyses and will be attached to the clinical study report.

2 STUDY OVERVIEW

2.1 STUDY DESIGN

This is a randomized, controlled, open-label, global Phase 3 study comparing overall survival following treatment with the anti-PD-1 monoclonal antibody tislelizumab to ICC given as a second line treatment in patients with advanced unresectable/metastatic ESCC that has progressed during or after first line therapy.

Before initiating this Phase 3 study in Japan, a sub-study investigating the safety, tolerability, PK and preliminary efficacy in Japanese patients is planned (see Appendix 13 of the Protocol Amendment 4 for details). The separate analysis plan will be planed for Japan sub-study.

After providing written informed consent, completing all screening assessments, and being confirmed as eligible for study participation, approximately 500 patients will be randomized 1:1 to receive either tislelizumab monotherapy or investigator chosen chemotherapy (paclitaxel/docetaxel/irinotecan). The choice of chemotherapy must be determined prior to randomization.

At randomization, patient enrollment will be stratified by the following 3 factors:

- Region (Asia [excluding Japan] vs Japan vs United States [US]/European Union [EU])
- Eastern Cooperative Oncology Group performance status (0 vs 1)
- ICC option (paclitaxel vs docetaxel vs irinotecan)

After randomization, patients will then begin open-label treatment with one of the following regimens.

Arm A: Tislelizumab 200 mg intravenously (IV), Day 1, given every 21 days

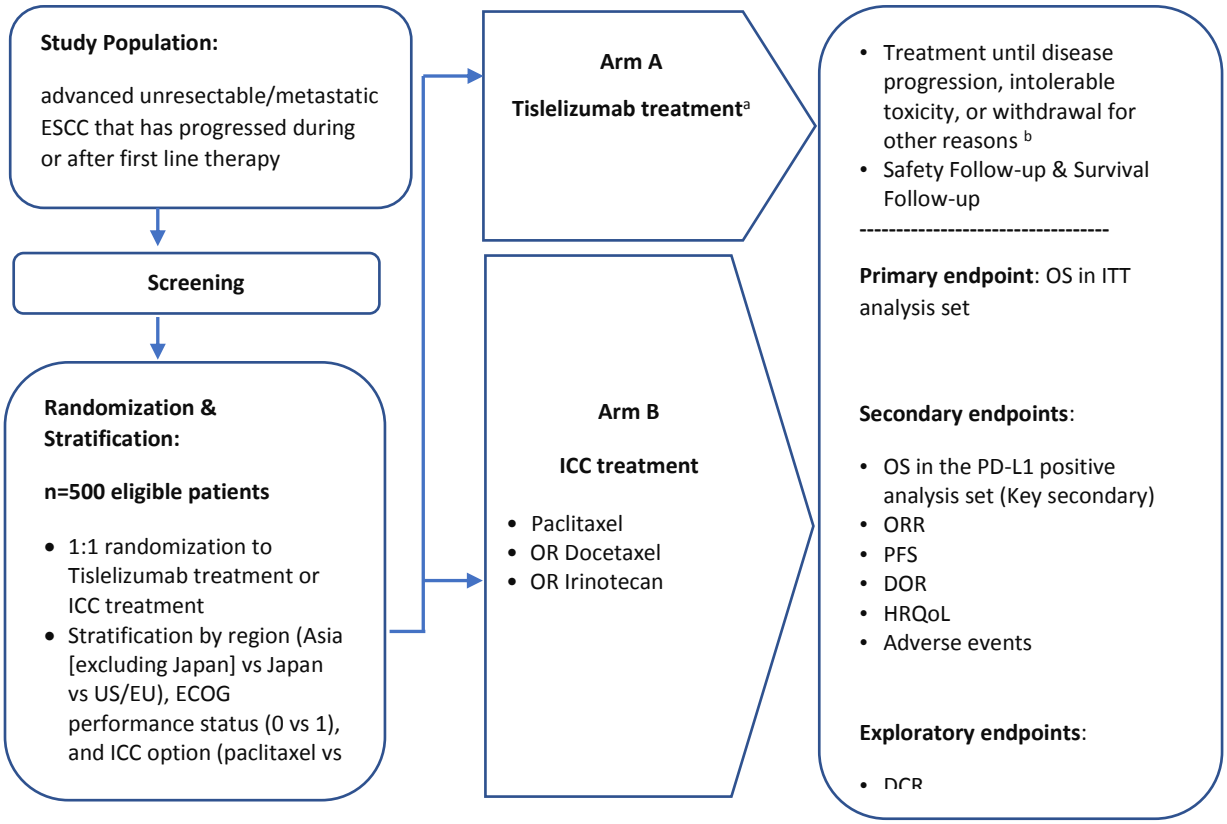
Arm B: One of the following three single-agent chemotherapies as determined by the investigator

- paclitaxel 135-175 mg/m² IV on Day 1, given every 21 days
  - Note, paclitaxel may also be given in doses of 80-100 mg/m^2 IV on a weekly schedule, according to local and/or country specific guidelines for standard of care
  - Japan: 100 mg/m^2 IV on Day 1, 8, 15, 22, 29, and 36, followed by one week of rest
- **OR** docetaxel 75 mg/m^2 IV on Day 1, given every 21 days
  - Japan: 70 mg/m^2 IV on Day 1, given every 21 days
- **OR** irinotecan 125 mg/m^2 IV on Days 1 and 8, given every 21 days

Cross-over between chemotherapy treatments or between chemotherapy and tislelizumab treatment arms during the study treatment period will not be allowed.

Study treatment will be administered until disease progression, intolerable toxicity, or another treatment discontinuation criterion is met. Tislelizumab treatment beyond initial investigator-assessed RECIST v1.1 defined progression is only permitted if the patient has evidence of "pseudo-progression" (Section 7.17.1 of the Protocol Amendment 4).

The study design schema is as follows:



Abbreviations: ECOG, Eastern Cooperative Oncology Group; ESCC, Esophageal squamous cell carcinoma; ICC, investigator chosen chemotherapy; ITT, Intention-to-treat; OS, Overall survival; ORR, Overall response rate; PD-L1, Programmed cell death protein ligand-1; PFS, Progression-free survival; DOR, Duration of response; HRQoL, Health Related Quality of Life; DCR, Disease control rate; ADA, Anti-drug antibody.

a. The initial infusion (Cycle 1, Day 1) will be administered over a period of 60 minutes. If this infusion is well tolerated, subsequent infusions may be administered over 30 minutes. After tislelizumab infusion, patients will be further monitored for at least 1 hour during Cycles 1 and 2. From Cycle 3 onward, a post-infusion monitoring period of at least 30 minutes will be required.

b. At the discretion of the Investigator, patients randomized to receive tislelizumab may be treated beyond progression under protocol defined conditions. See Section 7.17.1 of the Protocol Amendment 4.

2.2 STUDY ASSESSMENTS

Progression free survival and tumor response will be assessed by the Investigator using RECIST v1.1 criteria (Eisenhauer EA, 2009). Tumor imaging (CT with or without contrast or MRI) must be performed within 28 days prior to randomization. On-study tumor assessments will occur approximately every 6 weeks (±7 days) for 6 months, then every 9 weeks (±7 days) until disease progression. 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 a new anti-cancer therapy, disease progression, loss to follow up, withdrawal of consent, death, or until study termination, whichever occurs first.

Health related quality of life (HRQoL) will be collected via patient reported outcomes (PRO) instruments using the EORTC QLQ-C30, EORTC QLQ-OES18 and the EQ-5D-5L at baseline, Cycles 1-6 Day 1 or at the End of Treatment Visit (whichever occurs first), and at the Safety Follow-up Visit. Patients will be evaluated for any AEs and serious adverse events (SAEs) occurring up to 30 days after the last dose of study drug (all severity grades, per NCI-CTCAE v.4.03) or initiation of new anticancer therapy, whichever occurs first, and immune-related AEs (immune-mediated adverse events) occurring up to 90 days after the last dose of study drug regardless of initiation of a subsequent anticancer therapy. All drug-related SAEs will be recorded by the investigator after treatment discontinuation until patient death or loss to follow-up, whichever occurs first. All study drug related SAEs will be followed until they resolve to baseline or $\leqslant$ Grade 1, the Investigator assesses the AE as stable and unlikely to improve, or the patient is lost to follow-up, whichever occurs first.

Safety monitoring will be performed by an Independent Data Monitoring Committee (IDMC). The IDMC may recommend modifications to the study, including termination due to safety. The functions and membership of the IDMC will be described in an IDMC Charter.

3 STUDY OBJECTIVES

3.1 PRIMARY OBJECTIVES

To compare the overall survival (OS) in Intent-to-treat patients following treatment with tislelizumab vs. investigator chosen chemotherapy (ICC) when given as second line treatment in patients with advanced unresectable/metastatic Esophageal Squamous Cell Carcinoma (ESCC).

3.2 SECONDARY OBJECTIVES

3.2.1 Key Secondary Objective:

- To compare the OS in the PD-L1 positive population following treatment with tislelizumab versus investigator chosen chemotherapy (ICC).

3.2.2 Other Secondary Objectives:

- The following will be compared between the tislelizumab and the chemotherapy treatments based on assessment by investigator per RECIST v 1.1 criteria:
  - Overall response rate (ORR)
  - Progression-free survival (PFS)
  - Duration of response (DOR)
- To compare HRQoL endpoints between the tislelizumab and the chemotherapy treatment arms as assessed by patient reported outcome (PRO) measures: European EORTC QLQ-C30 index (EORTC QLQ-C30), the European esophageal cancer specific module QES 18 (EORTC QLQ-OES18), and the generic health state instrument Euroqol 5D (EQ-5D-5L).
- To compare the safety and tolerability between tislelizumab and the chemotherapy treatments

3.3 EXPLORATORY OBJECTIVES

- To characterize the disease control rate (DCR) with tislelizumab compared to chemotherapy
- To characterize the pharmacokinetics (PK) of tislelizumab
- To determine host immunogenicity to tislelizumab
- To explore potential predictive biomarkers (including but not limited to PD-L1 expression, gene expression profiling, tumor mutation burden, microsatellite instability (MSI), and tumor-infiltrated immune cells) and resistance mechanism immune cells

4 STUDY ENDPOINTS

4.1 PRIMARY ENDPOINTS

- Overall survival (OS) in the ITT analysis set – defined as the time from the date of randomization to the date of death due to any cause for all intent-to-treat patients.

4.2 SECONDARY ENDPOINTS

4.2.1 Key Secondary Endpoint:

- OS in the PD-L1 positive analysis set – defined as the time from the date of randomization until the date of death due to any cause for all PD-L1 positive patients.

4.2.2 Other Secondary Endpoints:

- ORR - defined as the proportion of patients who had complete response (CR) or partial

response (PR) assessed by the Investigators per RECIST v1.1

- PFS - defined as the time from the date of randomization to the date of first documentation of disease progression assessed by the Investigators per RECIST v1.1 or death, whichever occurs first
- DOR- measured from the time measurement criteria are first met for CR/PR (whichever is first recorded) until the first documentation of progression assessed by the Investigators per RECIST v1.1 or death, whichever comes first.
- HRQoL assessment of the subject's overall health status using European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire (QLQ)-C30 index, GHS, and the functional and symptoms scales, the EORTC QLQ esophageal cancer module OES18 index score and the symptoms scale scores, and the generic health state instrument Euroqol 5D (EQ-5D-5L)
- The incidence and severity of adverse events according to National Cancer Institute Common Terminology Criteria for Adverse Events (NCI-CTCAE) v4.03

4.3 EXPLORATORY ENDPOINTS

- Disease control rate (DCR)- defined as the proportion of patients who have CR, PR and stable disease (SD) assessed by the Investigators per RECIST v1.1
- Pharmacokinetic endpoints: summary of serum concentration of tislelizumab to include but not limited to trough serum concentration (C_{trough})
- Assessments of immunogenicity of tislelizumab to determine the incidence of anti-drug antibodies (ADA)
- Predictive biomarkers (including but not limited to PD-L1 expression, gene expression profiling, tumor mutation burden, MSI, and tumor-infiltrated immune cells) and resistance mechanism

5 SAMPLE SIZE CONSIDERATIONS

The sample size calculation is based on the primary efficacy analysis of OS comparison between tislelizumab and ICC arms in the ITT analysis set. Power of 82% and one-sided alpha of 0.025 are used in the sample size calculation. Assuming an OS-HR (Arm tislelizumab/Arm ICC) is 0.75 and a dropout rate of 5% per year, approximately 500 patients will be enrolled and randomized in a 1:1 ratio to Arms tislelizumab and ICC over 26-month period in order to accumulate approximately 400 deaths when median OS in the tislelizumab and ICC arms are 8 months and 6 months, respectively. The assumed OS-HR 0.75 is based on recently published results of anti PD-1 therapies in second line treatment of ESCC (Kojima et al 2019, Kato et al 2019, Huang et al 2019). .

6 STATISTICAL METHODS

6.1 ANALYSIS SETS

- Intent-to-Treat (ITT) analysis set – Includes all randomized patients. Patients will be analyzed according to their randomized treatment arm. This will be the primary analysis set for all efficacy analyses. Patients who were dead before the randomization but were inadvertently randomized into the study will be excluded from the ITT analysis set.
- The PD-L1 positive analysis set includes patients whose visually-estimated combined positive (vCPS) score ≥ 10% using VENTANA PD- L1 (SP263) CDx Assay. Visually-estimated combined positive is the total percentage of the tumor area covered by tumor cells with PD-L1 membrane staining and tumor- associated immune cells with PD-L1 staining at any intensity.
- Per-Protocol (PP) analysis set – Includes all randomized patients who received at least 1 dose of their assigned study drug and had no critical protocol deviations which are a significantly impact efficacy or safety evaluation. Critical protocol deviations are pre-specified in the protocol deviation specification. All cases of prospectively defined protocol deviations are identified prior to clinical database lock.
- Safety analysis set – Includes all randomized patients who received at least one dose of study drug. This will be the primary analysis set for all safety analyses. Patients will be analyzed by actual treatment received.
- The PK analysis set includes all patients who received at least one dose of tislelizumab per the protocol, for whom any post-dose PK data are available.
- The ADA analysis set includes all patients who received at least one dose of tislelizumab, have non-missing baseline ADA and at least one post-baseline ADA result.

6.2 DATA ANALYSIS GENERAL CONSIDERATIONS

6.2.1 Definitions and Computations

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

Baseline: For analysis in ITT analysis set and PD-L1 positive analysis set, a baseline value is defined as the last non-missing value collected on or before the randomization date. For analysis in safety analysis set, a baseline is defined as the last non-missing value collected on or before the first 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 a patient is still ongoing.

All calculations and analyses will be conducted using SAS version 9.2 or higher.

Conventions

Unless otherwise specified, all tabular outputs are summarized by tislelizumab, ICC options and total. 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 decimal.
- Age will be calculated as the integer part of (date of informed consent – date of birth + 1)/365.25
- P-values will be rounded to 4 decimal places. P-values that 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.
- Missing efficacy or safety data will not be imputed unless otherwise specified.
- For laboratory results collected as < or >, a numeric value, 0.0000000001 will be subtracted or added, respectively, to the value.
- For by-visit observed data analyses, percentages will be calculated based on the number of patients with non-missing data as the denominator, unless otherwise specified.
- For continuous endpoints, summary statistics will include n, mean, standard deviation, median and range (minimum and maximum).
- The unit of time duration is month unless otherwise specified.

6.2.2 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, prior/concomitant medications/procedures, and subsequent anti-cancer therapy are provided in Appendix 10.1, 10.2, 10.3, and 10.4

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

6.2.3 Adjustments for Covariates

The value of the stratification factors used at randomization (IWRS) ECOG performance status (0 vs 1) and ICC option (paclitaxel vs docetaxel vs irinotecan) will be used in stratified log-rank test and stratified Cox proportional hazard model for primary endpoint OS (ITT), key secondary endpoints OS(PD-L1 positive), and other secondary endpoints. The stratification factor region (Asia excluding Japan, Japan, US/EU) will not be adjusted in the model. However, subgroup analysis by region will be performed to assess the efficacy consistency by region

6.2.4 Multiplicity Adjustment

The primary endpoint of OS in the ITT analysis set will be tested once at a one-sided alpha of 0.025. If the null hypothesis for OS in ITT analysis set is rejected, the key secondary endpoint OS in PD-L1 positive analysis set will be tested.

6.2.5 Data Integrity

Before final statistical analysis begins, the integrity of the data should be reviewed to assure fit-for-purpose. The data set for analysis should be an accurate and complete representation of the subjects' relevant outcomes from the clinical database. All data should be complete and reviewed up to a pre-specified cutoff date. Consistency checks and appropriate source data verification should be complete.

6.3 SUBJECT CHARACTERISTICS

6.3.1 Subject Disposition

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

The number (percentage) of subjects randomized, treated, discontinued from treatment and discontinued from the study will be summarized. The primary reason for end of treatment (study drug discontinuation) and end of study will be summarized by categories in the eCRF. Study follow-up duration will be summarized descriptively.

Patient disposition and reasons for discontinuation will also be summarized by region (Asia and Europe/North America) for ITT analysis set.

6.3.2 Protocol Deviations

Protocol deviation criteria will be established together with its category/term of important or not important. Subjects 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.

Critical protocol deviation that significantly impacts efficacy or safety evaluation will be reviewed prior to data base lock according to the criteria defined in protocol deviation specification. The patient with those critical important protocol deviations will be excluded from per protocol analysis set.

Patient data listings of protocol deviation will be provided. A separate list of patients with critical protocol deviation will also be provided.

Important and non-important protocol deviations related to COVID-19 will be summarized.

6.3.3 Demographics and Other Baseline Disease Characteristics

Following demographics and other baseline characteristics collected will be summarized by using ITT analysis set and PD-L1 baseline status:

- Age (years) and age group (< 65 years, >= 65 years)
- Gender (Male, Female)
- Race
- Ethnicity
- Baseline ECOG
- Baseline height (cm)
- Baseline weight (kg)
- BMI (kg/m^2)
- PD-L1 status (vCPS score ≥ 10%, < 10 %, Missing)
- Smoking status (never, former, and current)
- Alcohol consumption (never, former, and current)

Demographics and other baseline characteristics will also be summarized by region (Asia and Europe/North America) for ITT analysis set.

The stratification factors (Region, ICC option, ECOG) collected in IWRS and CRF will also be summarized using ITT analysis set.

6.3.4 Disease History and Baseline Disease Symptoms

Following disease history and characteristics at study entry will be summarized using ITT analysis set and by PD-L1 baseline status:

- Number of Patients with Metastatic Diagnosis at Study Entry
- Time from Metastatic Diagnosis to Study Entry (month)
- Prior Systemic Therapy
- Primary Site of Esophageal Cancer
- Histologic Grade
- Target lesions sum of diameters by investigator
- Target lesion and non-target lesion location at Study Entry

Histology Grade Cancer associated symptoms at baseline will be coded by MedDra version 22.0 and summarized by system organ class, preferred term, and NCI CTCAE 4.03 grade using ITT analysis set.

Disease history and characteristics at study entry will also be summarized by region (Asia and Europe/North America) for ITT analysis set.

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

6.3.5 Prior Anti-Cancer Systemic Therapy, Radio Therapy and Surgeries

6.3.5.1 Prior Anti-Cancer Systemic Therapy

The number (percentage) of patient with at least one prior anti-cancer systemic therapy, maximal number of regimen of prior anti-cancer systemic therapies (0, 1, 2, >=3) and treatment setting (adjuvant/neoadjuvant, locally advanced, metastatic), duration of last prior anti-cancer systemic therapy (months) , and time from end of last anti-cancer systemic therapy to study entry (months) will be summarized by using ITT analysis set.

The prior anti-cancer systemic will also be summarized by Anatomical Therapeutic Chemical (ATC) class and preferred term based on treatment setting (Adjuvant, Neoadjuvant, Locally advanced, and Metastatic). In addition, the last prior anti-cancer systemic therapy before study entry will be summarized by ATC class and preferred term.

The prior anti-cancer systemic will also be summarized by region (Asia and Europe/North America) for ITT analysis set.

6.3.5.2 Prior Anti-Cancer Radio Therapy

The number (percentage) of patient with at least one prior anti-cancer radio therapy, treatment intent, treatment setting, time from last radio therapy to study entry (months), and treatment site (irradiated) will be summarized by using ITT analysis set.

6.3.5.3 Prior Anti-Cancer Surgeries

The number (percentage) of patient with at least one prior anti-cancer surgery, curative intent, will be summarized by using ITT analysis set.

Patient data listings of prior anti-cancer systematic therapy, radio therapy and surgeries will be provided.

6.3.6 Prior and Concomitant Medication and Therapy

Prior and concomitant medications will be assigned an 11-digit code using the World Health Organization Drug Dictionary drug codes. Concomitant medications will be further coded to the appropriate Anatomical Therapeutic Chemical code indicating therapeutic classification. Prior and concomitant medications will be summarized and listed by Anatomical Therapeutic Chemical (ATC) class and WHO drug preferred term for the safety analysis set.

Prior medications is defined as medications that stopped before the first dose of study drug.

Concomitant medications is defined as medications that (1) started before the first dose of study drug and were continuing at the time of the first dose of study drug, or (2) started on or after the date of the first dose of study drug up to 30 days after the patient's last dose. In addition, concomitant medication also includes medications associated with an immune-mediated adverse event recorded up to 90 days after last dose of study drug. Patients who received concomitant systemic corticosteroid will also be summarized.

Patient data listings of prior and concomitant medication will be provided

6.3.7 Subsequent Anti-Cancer Therapy

Subsequent anti-cancer therapy is defined as the anti-cancer therapy started after the last dose date of study drug. A summary of number and pecentage of patients who received subsequent systematic anticancer therapy/immunotherapy by arm will be provided based on ITT and PD-L1 positive analysis sets.

Patient data listings of subsequent anti-cancer therapy will be provided.

6.3.8 Medical History

Medical History will be coded using MedDRA (version 22.0 or later). The number (percentage) of subjects reporting a history of any medical condition, as recorded on the CRF, will be summarized by system organ class and preferred term for the ITT population.

Patient data listings of medical history will be provided.

6.4 EFFICACY ANALYSIS

The primary and key secondary endpoints of OS in the ITT and PD-L1 positive analysis sets will be tested sequentially at a one-sided alpha of 0.025. No statistical inference is planned for the other efficacy and safety analyses; the p-values from these analyses will be calculated for descriptive purpose only.

6.4.1 Primary Efficacy Endpoints

The primary efficacy endpoint overall survival (OS) measured from the date of randomization to date of death from any cause. In absence of death on or before data cut-off, OS will be censored either at the date that the patient was last known to be alive or the date of data cut-off, whichever comes earlier (Table 1).

Overall Survival (months) = (OS End Date – Date of Randomization + 1)/30.4275,

where OS End Date is defined below.

Table 1: Censoring Rules for Overall Survival

Situation	OS End Date	Censored
Death before/on cut-off	Date of death*	No
Death after cut-off	Date of data cut-off	Yes
Subject still on treatment at data cut-off	Date of data cut-off	Yes

Other	Date last known to be alive **	Yes

* 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 maximum of last available date showing patients alive +1 and first day of year/month of death date.

** The last know alive date is minimum of last available date showing patients alive and cutoff date. In general, any data after clinical cutoff date cannot be considered. The only exception is the data on the survival follow up CRF page after planned survival sweep with 100% clean and verified survival follow up data. After survival sweep, data collected from survival follow up page should be kept in the database even after clinical cutoff date. Any data that are used in last known alive date derivation need to be verified.

The null hypothesis to be tested is:

$$H_0: \text{OS in tislelizumab arm} = \text{OS in ICC arm}$$

against the alternative hypothesis:

$$H_1: \text{OS in tislelizumab arm} \neq \text{OS in ICC arm}$$

The primary analysis of overall survival will be analyzed in ITT once the targeted number of deaths (approximately 400) is reached. The p-value from one-sided log-rank test will be calculated, stratified by selected stratification factors at randomization: ECOG performance status (0 vs 1) and ICC option (paclitaxel vs docetaxel vs irinotecan). Median of OS with 95% confidence interval, if estimable, will be constructed with a generalized Brookmeyer and Crowley method. The cumulative probabilty of OS at 6 and 12 months, if estimable will be caculated and presented with two-sided 95% confidence interval by using Greenwood's formula.

The treatment effect will be estimated by fitting a Cox regression model to the OS times including treatment arm as a factor and stratified by ICC and ECOG performance status. From this model, the hazard ratio (HR) of OS will be estimated and presented with a two-sided 95% confidence interval.

Kaplan-Meier survival curve for each arm will also be provided.

6.4.1.1 Sensitivity Analysis

The following sensitivity analyses are planned:

- The analysis in primary efficacy will be repeated based on PP analysis set, PD-L1 positive analysis set, modified ITT analysis set. Modified Intent-to-Treat analysis set is defined as ITT analysis set excluding patients who are randomized but did not receive any dose of assigned study treatment.
- An unstratified analysis of OS will repeated in ITT analysis set. The treatment effect will be estimated by fitting a Cox regression model to OS times only including treatment arm as a factor.
- The treatment effect will be estimated by fitting a Cox regression model to the OS times

including treatment arm as a factor and stratified by ICC and ECOG performance status based on Case Report Form strata.

- The additonal PD-L1 categories are potentially planned applying different vCPS score cut-off.

If necessary, restricted mean survival time (RMST, Uno H, Claggett B, Tian L, Inoue E, et al. 2014), Max-combo (Satrajit R, Keaven A, Jiabu Y, Pralay M, 2019) will be performed to account for the possible non-proportional hazard effects.

In this study, patients can receive subsequnet systematic anticancer therapies (Immune checkpoint inhibitor or chemotherapy or target therapy) after treatment discontinuation which will casue bias in primary OS analysis. More sensitiviy analysis (i.e.,Rank Preserving Structural Failure Time (RPSFT) may be explored and is provided in Appendix 10.6.

6.4.2 Secondary Efficacy Endpoints

6.4.2.1 Overall Survival

The same statistical methods described above in the primary efficacy analysis will be applied for the key secondary efficacy endpoint (OS in the PD-L1 positive analysis set).

6.4.2.2 Objective Response Rate

The difference in ORR per RECIST v1.1 will be evaluated using Cochran-Mantel-Haenszel (CMH) test adjusting for ECOG performance status, and ICC option for the ITT analysis set and PD-L1 positive analysis set. The common odds ratio for ORR adjusted for strata and its two-sided 95% CIs will be calculated. ORR and difference in ORR as well as Clopper-Pearson 95% CI will be calculated.

Best overall response (BOR), defined as the best response recorded from randomization until data cut-off or the start of new anticancer treatment. Patients with no post-baseline response assessment (due to any reason) will be considered not assessable (NA) for BOR. The proportion for each of the response categories (e.g. CR, PR, SD, PD, NE, and NA) will be presented by treatment arm.

6.4.2.3 Progression-Free Survival

Progression-Free survival is defined in section 4.2. Progression-free survival censoring rule for PFS primary and sensitivity analysis are described in Table 2, which follows United States (US) Food and Drug Administration (FDA) Guidance for Industry, Clinical Trial Endpoints for Approval of Cancer drugs and Biologics (2018).

Table 2: The primary and secondary censoring rules for the derivation of PFS

No.	Situation	Date of Progression or Censoring	Primary Analysis	Sensitivity Analysis 1	Sensitivity Analysis 2
1	No baseline or any post-baseline tumor assessments and	Randomization date	Censored	Censored	Censored

	without death within two tumor assessments specified in protocol from reference start date				
2	Progression documented between scheduled visits	Date of first radiologic PD assessment	Progressed	Progressed	Progressed***
3	No progression at the time of data cut-off or withdrawal from study	Date of last adequate radiologic assessment prior to or on date of data cut-off or withdrawal from study	Censored	Censored	Censored
4	New anticancer treatment started	Date of last adequate radiologic assessment prior to or on date of new anticancer treatment	Censored	-	Censored
5	Death before first PD assessment	Date of death	Progressed	Progressed	Progressed***
6	Death between adequate assessment visits*	Date of death	Progressed	Progressed	Progressed***
7	Death or progression after more than one missed visit**	Date of last adequate radiologic assessment before missed tumor assessments	Censored	Censored	Progressed or died
8	No baseline or any post-baseline tumor assessments and died within two tumor assessments specified in protocol weeks from randomization date	Date of death	Progressed	Progressed	Progressed***

*Adequate tumor assessment is a radiologic assessment of CR, PR, SD, non-CR/non-PD or PD as determined by the investigators.

** More than one missed visit is defined if the duration between the last tumor assessment and death or PD is longer than D2. The D2 is defined as two times protocol

specified interval between tumor assessments (TAs) plus the protocol allowed window around the assessments. Since tumor assessment is scheduled as once every 6 weeks for first 6 months and once every 9 weeks afterwards with one-week window, D2 is 12 weeks + 1 week in the first 6 months and 18 weeks + 1 week afterwards.

*** Progression date for PFS event will be the earliest date of events defined in 2,4,5,6,8.

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 baseline tumor assessment (No. 1), the patient will be censored on the randomization date. However, if the patient died within two consecutive tumor assessments specified in protocol after randomization and did not receive new anticancer treatment, the date of death will be the PFS event date (not censored).
    b. If a patient had new anticancer treatment before PD or death (No. 4), 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 more than one assessment before PD or death (No. 7), the patient will be censored on the date of the last tumor assessment before PD or death. Not that if a patient is censored by both this criterion and the anticancer treatment criteria, the earliest censoring date will be used.
    d. Otherwise, if a patient has event (No. 2, No. 5, or 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. 3, No. 4).
3. In sensitivity analysis 1, the PFS event date will be derived ignoring new anti-cancer therapy.
4. In sensitivity analysis 2, any PD or death after more than one missing tumor assessment will be considered as a PFS event.

PFS primary analysis is based on ITT analysis set and PD-L1 positive analysis set. The sensitivity analysis is based on ITT analysis set only. The number and percentage of patients who were dead/progressed, censored with reason of censoring will be summarized for each arm. The p-value from one-sided log-rank test will be calculated, stratified by selected stratification factors at randomization: ECOG performance status (0 vs 1) and ICC option (paclitaxel vs docetaxel vs irinotecan). Median of PFS with 95% confidence interval, if estimable, will be constructed with a generalized Brookmeyer and Crowley method. The cumulative probabilty of PFS at 3 and 6 months, if estimable will be caculated and presented with two-sided 95% confidence interval by using Greenwood's formula.

The treatment effect will be estimated by fitting a Cox regression model to the PFS times including treatment arm as a factor and stratified by ICC and ECOG performance status. From this model, the hazard ratio (HR) of PFS will be estimated and presented with a two-sided 95%

confidence interval.

An unstratified analysis of PFS will repeated in ITT analysis set. The treatment effect will be estimated by fitting a Cox regression model to PFS times only including treatment arm as a factor.

Kaplan-Meier survival curve for each arm will also be provided.

Sensitivity Analysis for Progression-Free Survival

In order to evaluate the robustness of the PFS per RECIST 1.1, two sensitivity analyses with a different set of censoring rules will be performed (Table 2). The first sensitivity analysis is the same as the primary analysis except that it uses the actual reported date of progression or death to define PFS regardless of the use of new anti-cancer therapy. The second sensitivity analysis is the same as the primary analysis except that is uses the actual reported date of progression or death to define PFS regardless of the progression or death occur after more than one missed tumor assessment visit.

6.4.2.4 Duration of Response

Duration of Response (DOR) is defined in Section 4.2 as progression/death event free time counted from the first objective response date to the first documented radiological PD date/or death date, whichever occur first. All the censoring rules for PFS primary analysis (Table 2Table 2). should be applied to DOR as well. DOR analysis is based on ITT analysis set and PD-L1 positive analysis set.

The p-value from one-sided unstratified log-rank test will be calculated for descriptive purpose only. Hazard ratio will be based on a unstratified Cox regression model only including treatment as covariate. Median of DOR with 95% confidence interval, if estimable, will be constructed with a generalized Brookmeyer and Crowley method. The cumulative probabilty of DOR at 3 and 6 months, if estimable will be caculated and presented with two-sided 95% confidence interval by using Greenwood's formula.

Kaplan-Meier survival curve for each arm will also be provided.

6.4.2.5 Health-Related Quality of Life (HRQoL)

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

The EORTC-QLQ-OES18 (Wen Y, et al., 2015) is the specific esophageal symptoms module of the QLQ-C30, and includes 18 questions: 6 single item subscales measuring saliva swallowing,

choking, dry mouth, taste, coughing, and talking. It also includes 12 items grouped into 4 subscales: dysphagia (3 items), eating (4 items), reflux (2 items), and pain (3 items).

The EQ-5D-5L comprises a descriptive module and an EQ Visual Analogue scale (EQ VAS). The EQ-5D-5L descriptive module comprises the following 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems, and extreme problems. The EQ VAS records the respondent's self-rated health on a 0 to 100 scale, with 100 labelled 'the best health you can imagine' and 0 'the worst health you can imagine'. Lower scores in descriptive dimension indicates better HRQoL and higher VAS score indicates better health state.

EORTC Scoring Derivation

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

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

Raw Score (RS)

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

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

Derived Scale (DS)

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

For functional scales:

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

For symptom scales and global health status:

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

Refer Table 3 and Table 4 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) ÷ 10

C30 index- score=∑ [(100-DS of Physical functioning, 100-DS of Role functioning, 100-DS of Emotional functioning, 100-DS of Cognitive functioning, 100-DS of Social functioning, 100-DS of global QOL, DS of Fatigue, Nausea, Vomiting, Pain, Dyspnoea, Insomnia, Appetite loss, Constipation, Diarrhea, Financial Difficulty] ÷ 15

Table 3: Scoring of QLQ-C30

	Scale	Number of items	Item range	Item Numbers
Global health status/ QoL Global health status/QOL	QL2	2	6	29,30
Functional Scales				
Physical functioning	PF2	5	3	1, 2, 3, 4, 5
Role functioning	RF2	2	3	6, 7
Emotional functioning	EF	4	3	21, 22, 23, 24
Cognitive functioning	CF	2	3	20, 25
Social functioning	SF	2	3	26, 27
Symptom Scales				
Fatigue	FA	3	3	10, 12, 18
Nausea and vomiting	NV	2	3	14, 15
Pain	PA	2	3	9, 19
Single Items				
Dyspnoea	DY	1	3	8
Insomnia	SL	1	3	11
Appetite loss	AP	1	3	13
Constipation	CO	1	3	16
Diarrhea	DI	1	3	17
Financial Difficulties	FI	1	3	28

Table 4: Scoring of QLQ-OES18

	Scale	Number of items	Item range	Item Numbers
Symptom Scales				
Dysphagia	DY	3	3	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.

All HRQoL measures will be summarized in ITT analysis set and PD-L1 positive analysis set.

Completion rates for the EORTC-QLQ-C30, QLQ-OES18, and EQ-5D-5L will be summarized separately at each visit. A questionnaire module is considered complete if at least one question is answered. In addition, the adjusted completion rate which defined as number of patients complete all questions divided by the number of patients still on study at relevant visit will also be summarized.

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 index score and derived score (functional scales/symptom scales/single items and the global scale) of EORTC-QLQ-C30, EORTC-QLQ-OES18 will be summarized as well as change from baseline using descriptive statistics.

For EQ-5D-5L, descriptive system will be summarized by visit and dimension in an ordinal scale using descriptive statistics. The EQ VAS and change from baseline will be summarized by visit in a continuous scale using descriptive statistics.

Time to clinically meaningful worsening in a HRQoL domains (for functional scales and global health status/) is defined as the time from randomization to the first time the difference between the current derived score and baseline was ≥ 10 in the worsening direction. A deterioration is not counted as an event if a subsequent improvement returned the overall worsening from baseline to less than 10 points. Patients without clinically meaningful worsening will be censored at the last time the HRQoL domain was assessed. Time to clinically meaningful worsening will be analyzed for comparing the difference between the two treatment arms using Cox model for selected domains, hazard ratio and its 95% confidence interval will be provided, and a forest plot for selected domains will be provided. Time to clinically worsening for a selected number of domains of major interest will be tested for comparing the difference between tislelizumab and ICC arms. The domains of major interest include: the global scale and physical function scale of QLQ-C30 and four symptom scale (dysphagia, eating, reflux, pain) from QLQ-OES18.

KM estimates by treatment arm, the hazard ratio estimates and their 95% CI will be provided. Hazard ratio is based on a Cox regression model including treatment as covariate and stratified by ECOG status and ICC options.

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

6.4.3 Exploratory Efficacy Endpoints

Disease control rate (DCR) defined as the proportion of patients who have CR, PR and SD assessed by the investigator per RECIST 1.1. DCR will be analyzed similarly to ORR analysis by using ITT analysis set and PD-L1 positive analysis set.

Waterfall plots will be provided for the maximum tumor shrinkage based on target lesion. The maximum tumor shrinkage based on target lesion used in the plots will be listed. The post-baseline nadir will be summarized using descriptive statistics. These analyses will be performed based on ITT analysis set and PD-L1 positive analysis set.

6.4.4 Subgroup Analyses

The following exploratory subgroup analyses may be conducted on the primary efficacy endpoint for OS in ITT analysis set.

- ICC options
- Geographic region
  - Asia vs Europe/North America
  - Asia (excluding Japan) vs. Japan vs. Europe/North America
- ECOG Performance status (0/1)
- Age group (< 65 years, >= 65 years)
- Gender
- Smoking status (former/current smoker, non-smoker)
- Race (White, , Other)
- Baseline PD-L1 expression category:
  - vCPS score >= 10%, vCPS score < 10%, missing

Geographic region (Asia vs Europe/North America) and Race (White vs Other) will also be summarized for the OS in PD-L1 positive analysis set.

KM estimates by treatment arm, the unstratified hazard ratio estimates and their 95% CI will be provided. The treatment effect will be estimated by fitting a Cox regression model to OS times only including treatment arm as a factor.

Country-specific subgroups may also be summarized per local regulatory requirements. For the secondary endpoints, PFS, ORR and DOR will be summarized by region (Asia vs Europe/North America) in ITT analysis set.

6.5 SAFETY ANALYSES

Safety will be assessed by the monitoring and recording of all AEs graded by NCI-CTCAE v4.03. Laboratory values (e.g., hematology, clinical chemistry, coagulation and urinalysis), vital signs, ECGs, and physician examinations will also be evaluated in defining the safety profile of each treatment arm. 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 analysis will be performed based on actual treatment received.

Treatment arm for safety analyses is defined as follows.

If a patient was randomized to the tislelizumab treatment arm, and received tislelizumab at least once, the treatment arm for this patient is tislelizumab; otherwise, the actual treatment arm for this patient is ICC, as determined by his/her first study drug received.

If a patient was randomized to the ICC treatment arm, and received assigned ICC treatment at least once, then the treatment arm for this patient is ICC. Otherwise, the actual treatment arm is determined by his/her first ICC drug received. If the patient only received tislelizumab, then the treatment arm for this patient is tislelizumab.

6.5.1 Extent of Exposure

Extent of exposure will be summarized by treatment arm. For ICC arm, extent of exposure will be summarized by study drug.

Extent of exposure will be summarized descriptively as the number of cycles received, duration of exposure (days), cumulative total dose received per patient, dose intensity and relative dose intensity.

The number (percentage) of patients requiring infusion related dose omission, dose interruption, dose delay, and dose reduction (in the ICC arm only) due to AEs will be summarized for each study drug. Consecutive dose omission/skips will only be counted once. Frequency of the above dose adjustments will be summarized by category.

- 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 drug.
- Cumulative total dose per subject will be computed as the sum of all of the doses received in each cycle for each study drug. The dose unit is mg for tislelizumab and mg/m^2 for ICC arm.
- Dose intensity is defined as total dose / duration of exposure (day) * cycle length in days. Relative dose intensity is defined as Dose intensity / planned dose intensity. The planned dose intensity is the total planned dose in a cycle. For a patient in the tislelizumab arm, the planned dose intensity is 200 mg Q3W. For ICC, use the plan dose capture in the case report form.

Treatment duration

- For the tislelizumab arm, Treatment duration = date of the last dose – date of first dose + 21.
- For patients taking Paclitaxel on a once weekly schedule,
  Treatment duration = date of the last dose – date of first dose + 7.
- For patients taking paclitaxel on a Q3W schedule,
  Treatment duration = date of the last dose – date of first dose + 21.
- For Japanese patients taking paclitaxel on Days 1, 8, 15, 22, 29, and 36, followed by one week of rest,

  Treatment duration = date of the last dose – date of first dose + 14.
- For patients taking docetaxel,
  Treatment duration = date of the last dose – date of first dose + 21.
- For patients taking irinotecan,
  Treatment duration = date of the last dose – date of first dose + 14.

6.5.2 Adverse Events

The AE verbatim descriptions (Investigator's description from the eCRF) will be classified into standardized medical terminology using Medical Dictionary for Regulatory Activities (MedDRA). Adverse events will be coded to MedDRA (Version 22.0 or higher) lower level term closest to the verbatim term. The linked MedDRA System Organ Class (SOC) and Preferred Term (PT) are also classified. All adverse event summaries are based on safety analysis set.

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 study drug up to 30 days following study drug discontinuation or initiation of new anti-cancer therapy, whichever occurs first. TEAEs also include all immune-mediated adverse events up to 90 days after the last dose of study drug regardless of whether or not the patient starts a new anticancer therapy. Only those AEs that were treatment emergent will be included in summary tables.

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

6.5.2.1 Treatment Emergent Adverse Event

An overall summary of TEAEs will summarize the number (%) of patients with

- At least one TEAE
- At least one TEAE with NCI-CTCAE grade ≥3
- At least one treatment-related TEAE
- At least one serious TEAE
- At least one TEAE leading to death
- At least one TEAE leading to discontinuation of study drug
- At least one TEAE leading to dose modification of study drug

- At least one immune-mediated adverse event
- At least one immune-mediated adverse event with NCI-CTCAE grade ≥3
- At least one infusion-related reaction
- At least one infusion-related reaction with NCI-CTCAE grade ≥3

Summaries of the following TEAEs will be provided:

- All TEAEs
  - All TEAEs by SOC
  - All TEAEs by SOC and PT
  - All TEAEs by SOC and PT by Region
  - Most frequently reported (incidence ≥ 5% in any treatment arm) TEAEs by SOC and PT
  - Treatment-related TEAEs by SOC
  - Treatment-related TEAEs by SOC and PT
  - Treatment-related TEAEs by SOC and PT by Region
  - Most frequently reported (incidence ≥ 5% in any treatment arm) Treatment-related TEAE by SOC and PT
- Serious TEAEs by SOC and PT
  - Most frequently reported (incidence ≥ 1% in any treatment arm) serious TEAE by SOC and PT
  - Treatment-related Serious TEAE by SOC and PT
  - Most frequently reported (incidence ≥ 1% in any treatment arm) Treatment-related serious TEAE by SOC and PT
- TEAEs with NCI-CTCAE grade ≥3 by SOC and PT
  - TEAEs with NCI-CTCAE grade ≥3 by SOC and PT by Region
  - Most frequently reported (incidence ≥ 1% in any treatment arm) TEAE with NCI-CTCAE grade ≥ 3 by SOC and PT
  - Treatment-related TEAE with NCI-CTCAE grade ≥ 3 by SOC and PT
  - Treatment-related TEAE with NCI-CTCAE grade ≥ 3 by SOC and PT by Region
  - Most frequently reported (incidence ≥ 1% in any treatment arm) Treatment-related TEAE with NCI-CTCAE grade ≥ 3 by SOC and PT
  - TEAEs leading to death by SOC and PT
  - Treatment-related TEAE Leading to Death by SOC and PT

- TEAEs Leading to Death Occurring in ⩾ 2 Patients by System Organ Class and Preferred Term

- TEAEs leading to treatment discontinuation by SOC and PT
  - Treatment-related TEAE Leading to Treatment Discontinuation by SOC and PT
- TEAEs leading to dose modification by SOC and PT
  - Treatment-related TEAE Leading to Dose Modification by SOC and PT

The types of Dose modification include dose held, dose interruption and dose reduction for ICC arm; dose held and dose interruption for tislelizumab arm.

In addition, different cutoff of AE summary may be provided.

6.5.2.2 Immune-related Adverse Event

Immune-mediated adverse events are of special interest and summarized by category within a pre-defined list in Appendix 10.5.

For immune-mediated adverse events, a summary of incidence based on the number of patients dosed or within the immune-mediated adverse event follow up period will be presented by category in the descending order of incidence based on the tislelizumab column.

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

- Immune-mediated adverse events by category
- Immune-mediated adverse events by category by region
- Immune-mediated adverse events with NCI-CTCAE grade ≥3 by category
- Immune-mediated adverse events with NCI-CTCAE grade ≥3 by category by region
- Immune-mediated adverse events by category and worst grade
- Immune-mediated adverse events leading to treatment discontinuation by category
- Immune-mediated adverse events leading to death by category
- Immune-mediated adverse events leading to dose modification by category

6.5.2.3 Infusion-related Adverse Event

For IRRs, a summary of incidence by SOC, PT and maximum severity will be provided, sorted by descending order of incidence within each SOC and PT based on tislelizumab column. Summaries of IRRs, IRRs with NCI-CTCAE grade ≥3, IRRs leading to treatment discontinuation, and IRRs leading to dose modification will also be provided by PT only, in descending order.

6.5.3 Death

Number and causes of deaths, classified by deaths within 30 days of last dose of study drug and deaths more than 30 days after the last dose, will be summarized based on ITT analysis set.

Patient data listings of death and reason will be provided.

6.5.4 Laboratory Values

Hematology, serum chemistry, thyroid function, will be summarized/listed for selected parameters described in Table 5. The coagulation, urinalysis, HBV/HCV serology, and pregnancy test results will be listed only.

Laboratory results will be summarized/listed using Système International (SI) units, as appropriate. For all quantitative parameters listed in Table 5, **Error! Reference source not found.**the actual value and the change from baseline will be summarized by visit and worst post-baseline visit using descriptive statistics. Plots of laboratory values/change from baseline over time will be provided for selected lab parameters.

Laboratory parameters are also graded according to CTCAE v4.03 and will be summarized by shifts from baseline CTCAE grades to maximum post-baseline grades. Laboratory parameters (e.g., glucose, potassium, sodium) with CTCAE grading in both high and low directions will be summarized separately. Shift tables will be used to summarize the grade change from baseline to worst post baseline value with counts and percentages for hematology and serum chemistry. The lab parameters with grades increased in more than 2 from baseline to worst post baseline will also be summarized.

Table 5：Clinical Laboratory Assessments

Serum Chemistry	Hematology	Coagulation	Urinalysis	Thyroid function	HBV and HCV serology
glucose, blood urea nitrogen [BUN] [or serum urea], creatinine, sodium, potassium, chloride, phosphorus, total bilirubin, ALT, AST, alkaline phosphatase, lactate dehydrogenase [LDH] [optional], total protein, albumin, creatine kinase (CK) and creatine kinase-cardiac muscle isoenzyme (CK-MB)	complete blood count [CBC], including red blood cell [RBC] count, hemoglobin, hematocrit, white blood cell [WBC] count with automated differential [neutrophils, eosinophils, lymphocytes, monocytes, and basophils], and platelet count	Prothrombin time (PT) and International Normalized Ratio (INR)	complete [including, but not limited to glucose, protein, ketones, and blood] and/or microscopic, if clinically indicated	TSH Free T3 Free T4	HBsAg, HBcAb and HCV antibody

6.5.5 Vital Signs

Descriptive statistics for vital sign parameters (systolic and diastolic blood pressure [BP], pulse rate, temperature, and weight) and changes from baseline will be presented by visit. For tislelizumab, the change from post-dose (end of infusion) to pre-dose also need to be summarized for all vital sign parameters except for weight. Vital signs will be listed by patients and visits. Descriptive statistics for vital sign parameters.

6.5.6 Ophthalmology Examination

Ophthalmology examination findings will be listed by patients and visits.

6.5.7 Electrocardiograms (ECG)

12-lead ECG recordings are required at Screening, End of Treatment and as clinically indicated. Patient listing of ECG will be provided for all ECG recordings.

The actual value and the change from baseline for QTc intervals will be summarized by visit and treatment arm using descriptive statistics.

Abnormal post-baseline QTcF results will be summarized with the following categories: increase of >30 msec, increase of > 60 msec, value of > 450 msec, value of > 480 msec, value of > 500 msec for each visit by treatment arm.

6.5.8 ECOG

A shift table from baseline to worst post baseline will be provided. Patient listing of ECOG performance score comparison between IRT and CRF will be provided for all ECOG recordings.

6.6 PHARMACOKINETIC ANALYSES

Pharmacokinetic samples were collected in this study as outlined in Appendix 1 in Protocol Amendment v4.0 and only from patients randomized to receive tislelizumab and in sites that are able to adequately perform sampling, handling and processing procedures outlined in the laboratory manual.

Tislelizumab (as well as by region and race) serum concentration data,including but not limited to C_{trough}, will be tabulated and summarized for each cycle at which pharmacokinetics are collected. Descriptive statistics will include means, medians, ranges, standard deviations and coefficient of variation (CV), and geometric mean, geometric CV as appropriate.

Additional PK analyses may be conducted as appropriate.

6.7 IMMUNOGENICITY

Samples to assess anti tislelizumab antibodies will be collected only in patients randomized to receive tislelizumab and in sites that are able to adequately perform sampling, handling and processing procedures outlined in the laboratory manual.

ADA attributes:

- **Treatment boosted ADA** is defined as ADA positive at baseline that was boosted to a 4-fold or higher-level following drug administration.

- **Treatment-emergent** ADA is defined as ADA negative at baseline and ADA positive post-baseline.
- **Transient ADA response** is defined as Treatment-emergent ADA detected only at 1 time point during treatment or follow-up, excluding last time point; or detected at 2 or more time points during treatment or follow-up, where the first and last positive samples are separated by less than 16 weeks and the last time point is negative.
- **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 only in the last time point or at a time point less than 16 weeks before a negative last sample.
- **Neutralizing ADA** is defined as ADA that inhibits or reduces the pharmacological activity.

ADA response endpoints:

- **ADA incidence** is defined as sum of 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 allow.

6.8 OTHER ANALYSIS

Distribution of PD-L1 expression, gene expression profile, tumor mutation burden (tumor tissue and/or blood), MSI and tumor-infiltrated immune cells may be examined in the ITT analysis set. Same analysis may be examined in on-progression samples if any collected. BOR, ORR, and PFS may be present by PD-L1 category. Other potential predictive markers may be assessed with the same approach.

Detailed information about biomarker analysis for exploratory endpoints will be in separate statistical analysis document and will not be described in the clinical study report.

The impact of COVID-19 to patient disposition, adverse events, exposure with dose modification, protocol violation/deviation will be summarized.

7 INTERIM ANALYSIS

There is no planned interim analysis for this study.

8 CHANGES IN THE PLANNED ANALYSIS

If the SAP needs to be revised after the study starts, 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.

9 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.

Eisenhauer EA, et al. (2009). New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer, 45(2), 228-47.

Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A, on behalf of the EORTC Quality of Life Group. The EORTC QLQ-C30 Scoring Manual (3rd Edition). Published by European Organisation for Research and Treatment of Cancer, Brussels 2001.

Hwang IK, Shih WJ, and DeCani JS. Group sequential designs using a family of type 1 error probability spending functions. Statistics in Medicine, 9:1439–1445, 1990.

National Comprehensive Cancer Network (NCCN) Guidelines Version 1. (2017). Esophageal and Esophagogastric Junction Cancers. Retrieved from

https://www.nccn.org/professionals/physician_gls/f_guidelines.asp#esophageal

U.S. Food and Drug Administration Cencer for Drug Evaluation Research (CDER) and Center for Biologics Evaluation and Research. (2007). Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics.

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

10 APPENDIX

10.1 IMPUTE PARTIAL DATES FOR CONCOMITANT MEDICATION

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

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

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

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

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

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

10.2 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 > death date, then set to death 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 >

death date, then set to death date.

10.3 IMPUTE PARTIAL DATES FOR SUBSEQUENT ANTI-CANCER SURGERY/PROCEDURE

When the start date of subsequent anti-cancer therapy is partially missing, the following rules will be applied to impute partial dates.

If start 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 end date > death date, then set to death date

If year of the start date is missing, do not impute. If imputed start date is after study discontinuation date, then set to study discontinuation date.

10.4 IMPUTE PARTIAL DATES FOR PRIOR ANTI-CANCER THERAPY (DRUG, SURGERY/PROCEDURE, RADIOTHERAPY)

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

- If start date of a disease history or prior therapy 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 > first dose date then set to first dose date – 1

If end date of a disease history or prior 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 end date > first dose date, then set to first dose date -1

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. If imputed start date/end date is after randomization date - 14 , then set to randomization date - 14.

10.5 IMMUNE-MEDIATED ADVERSE EVENT CATEGORY LIST

Category
Immune-mediated adrenal insufficiency
Immune-mediated anaemia
Immune-mediated colitis
Immune-mediated hepatitis
Immune-mediated hyperthyroidism
Immune-mediated hypothyroidism
Immune-mediated myocarditis
Immune-mediated myositis/rhabdomyolysis
Immune-mediated nephritis and renal dysfunction
Immune-mediated nervous system disorder
Immune-mediated ocular disorder
Immune-mediated pancreatitis
Immune-mediated pituitary dysfunction
Immune-mediated pneumonitis
Immune-mediated skin adverse reaction
Immune-mediated thrombocytopenia
Immune-mediated thyroiditis
Immune-mediated type 1 diabetes mellitus
Other immune-mediated reactions

10.6 SENSITIVITY ANALYSES FOR SUBSEQUENT THERAPY

Rank Perserving Structural Failure Time (RPSFT) survival analysis model of OS accounting for the use of subsequent anticancer therpy are planned.

Due to there is no standard of care for 3rd line ESCC patients, we assume the subsequent immune checkpoint inhibitors contribute most effect for OS than other subsequent anti-cancer therapy. Under this assumption, we only focus on adjustment of subsequent immune checkpoint inhibitors by using RPFST model. RPSFT assumes 'common treatment effect' assumption which implies that the treatment effect of tisleilizumab and other immune checkpoint inhibitors (ICI) are identical and the treatment effect is indedpendent of exposure timing in the trial. uses a counterfactual framework to estimate the causal effect of immune checkpoint inhibitors, where counterfactual surival times refer to those that would have been observed if no subsequent immune checkpoint inhibitors had been given. It is assumed that counterfactual survival times are independent of treatment arm and grid method (g-estimation) is used to determine a value for the treatment effect. The p-value, median OS, hazard ratio will be constructed based on counterfactual dataset. The symmetrical test-based 95% confidence intervals were obtabined by inflating the standard error of the log-hazard ratio to preserve the ITT p-value.

The respective sandard error of the estimated hazard ratio of counterfactural dataset:

$$SE = \frac{|\log\widehat{(HR_c)}|}{Z_{1-p/2}}$$

The symmetrical test-based 95% confidence intervals:

$$\exp(\log\widehat{(HR_c)} \pm Z_{1-\frac{\alpha}{2}} \times SE)$$