STATISTICAL ANALYSIS PLAN

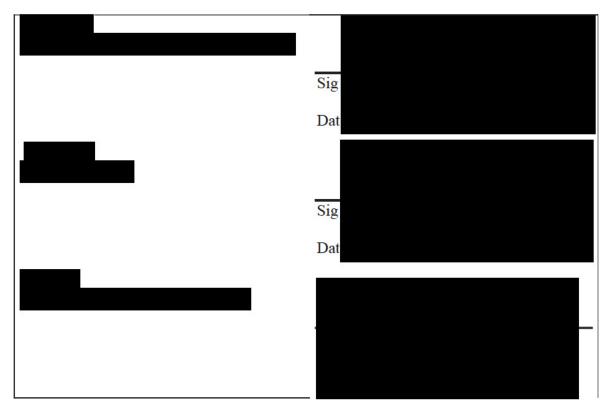|	
Study Protocol Number:	BGB-A317-306
Study Protocol Title:	A Randomized, Placebo-Controlled, Double-Blind Phase 3 Study to Evaluate the Efficacy and Safety of Tislelizumab (BGB-A317) in Combination with Chemotherapy as First-Line Treatment in Patients with Unresectable, Locally Advanced Recurrent or Metastatic Esophageal Squamous Cell Carcinoma
Date:	Feb 18, 2022
Version:	Version 2.0

VV-CLIN-092227   Version   2.0

SIGNATURE PAGE

Author:

Sign

Dat

Approval

Sig

Dat

Sig

Dat

TABLE OF CONTENTS

VV-CLIN-092227   Version 2.0

VV-CLIN-092227   Version   2.0

Tracking Changes

Update Content	Update Impact	Section (Current index)
Cutoff -> cutoff Change TAP score to PD-L1 score Change gender to sex	minor	overall
Update prior definitive therapy criteria for clarity	minor	2.1
Change biomarker wording	minor	3.2
Clarity added for estimand Clarity added to "Asia": i.e., including Asia (excluding Japan" and Japan as described in the protocol	minor	4 & overall
Add clarity to safety analysis set	minor	7.1
For prior definitive therapy, add cancer-related to surgery Update summary items	minor	2.1 7.3.5
drop by indication table	minor	7.3.6
Add clarity for sensitivity analysis, drop sensitivity analysis 3(unpooled region as stratification factor), sensitivity analysis 4(PD-L1 status as stratification factor) for post-hoc as needed	minor	7.4.1
Estimated using Greenwood's formula update reference to Kalbfleisch and Prentice 1980 Drop rank preserving and two-stage supportive analysis	minor	7.4.2
Clarify that Tradition Chinese Medicine are not considered as next line anticancer therapy for efficacy and safety	minor	7.4.3
adding ND from IRC as valid assessment of PFS Add more clarity to PFS2 censoring	minor	7.4.4
Subsequent anti-cancer therapy also summarized by region, adding Asia(excluding Japan) vs.	minor	7.4.6

VV-CLIN-092227   Version   2.0

Japan vs. Rest of world for OS by subgroup analysis		
Add infusion rate decreased into the analyses of dose modification Add tislelizumab/placebo related TEAE leading to death occurring >=2 pts summary Add rationale of TEAE definition change Update AE summary Add clarity of exposure adjusted AE	major	7.5 & 7.5.1
Add lab parameter Add by region tables	minor	7.5.3
Update AE partial missing imputation	minor	11.2

VV-CLIN-092227   Version   2.0

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
BCLC	Barcelona Clinic Liver Cancer
BGB-A317	Code name for monoclonal antibody tislelizumab
BID	Twice daily
BIRC	Blinded Independent Review Committee
BOR	Best overall response
CBC	Complete blood count
CBR	Clinical benefit rate
CI	Confidence interval
C_{max}	Maximum observed plasma concentration
C_{min}	Minimum observed plasma concentration
CP	Child-Pugh
CR	Complete response
CSR	Clinical Study Report
CT	Computed tomography
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
EQ-5D	European Quality of Life 5-Dimensions
EORTC QLQ-HCC 18	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Hepatocellular Carcinoma 18 Questions
EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30
FDA	Food and Drug Administration
FDG-PET	Fluorodeoxyglucose-position emission tomography
FFPE	Formalin-fixed paraffin-embedded
GCP	Good Clinical Practice
HbcAb	Hepatitis B core antibody
HbsAg	Hepatitis B surface antigen
HBV	Hepatitis B virus
HCC	Hepatocellular carcinoma
HCV	Hepatitis C virus
HR	Hazard ratio
HRQoL	Health-related quality of life
ICF	Informed consent form
ICH	International Conference on Harmonisation
INR	International Normalized Ratio
imAE	Immune-mediated adverse event
ITT	Intent-to-Treat

VV-CLIN-092227   Version 2.0

IV	Intravenous(ly)
MedDRA	Medical Dictionary for Regulatory Activities
MRI	Magnetic resonance imaging
MTD	Maximum tolerated dose
NCI-CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
ORR	Objective response rate
OS	Overall survival
PD	Progressive disease
PD-1	Programmed cell death protein-1
PFS	Progression-free survival
PK	Pharmacokinetic(s)
PO	Orally
PP	Per-Protocol
PR	Partial response
ROW	Rest of world
Q2W	Once every 2 weeks
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
TACE	Transarterial chemoembolization

VV-CLIN-092227   Version 2.0

TEAE	Treatment-emergent adverse event
ULN	Upper limit of normal

VV-CLIN-092227   Version   2.0

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 A317-306: A Randomized, Placebo-Controlled, Double-Blind Phase 3 Study to Evaluate the Efficacy and Safety of Tislelizumab (BGB-A317) in Combination with Chemotherapy as First-Line Treatment in Patients with Unresectable, Locally Advanced Recurrent or Metastatic Esophageal Squamous Cell Carcinoma. The focus of this SAP is for the planned interim analysis and the final analysis specified in the study protocol. This SAP is based on BGB-A317-306 Protocol Amendment 4.0 dated on 2021.

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.

2   STUDY OVERVIEW

2.1   STUDY DESIGN

This is a randomized, placebo-controlled, double-blind, global Phase 3 study comparing efficacy and safety following treatment with the anti-programmed cell death protein-1 (PD-1) monoclonal antibody tislelizumab in combination with standard chemotherapy compared to placebo in combination with chemotherapy when given as the first-line treatment in patients with unresectable, locally advanced recurrent or metastatic ESCC.

The study design schema is as follows:

```
┌─────────────────────────┐
│ Study Population:       │          ┌──────────────────┐        • Treatment until disease
│ unresectable, locally   │          │                  │          progression, intolerable
│ advanced recurrent or   │          │     Arm A        │          toxicity, or withdrawal for
│ metastatic ESCC         │          │ Tislelizumab in  │          other reasons
└─────────────────────────┘          │ combination with │        • Safety Follow-up & Survival
              │                      │ chemotherapy     │          Follow-up
              ▼                      │                  │        --------------------------------
┌─────────────────────────┐          └──────────────────┘        Primary endpoint:
│      Screening          │                                      • OS
└─────────────────────────┘
              │                                                  Secondary endpoints:
              ▼                                                  • PFS, ORR, and DOR assessed
┌─────────────────────────┐                                        by investigator
│ Randomization &         │                                      • OS in PD-L1 vCPS ⩾ 10%
│ Stratification:         │                                        subgroup
│                         │                                      • HRQoL
│ n=622 eligible patients │                                      • Adverse events
│ • 1:1 randomization to  │
│   tislelizumab+chemo    │                                      Exploratory endpoints:
│   or placebo+chemo      │          ┌──────────────────┐        • DCR
│   Stratification        │          │                  │        • PFS, ORR, DOR, and DCR
│ • Geographic region     │          │     Arm B         │          assessed by BIRC
│   (Asia [excluding      │          │ Placebo in combination     • PFS2
│   Japan] vs Japan vs    │          │ with chemotherapy│        • Correlation of biomarker
│   Rest of World)        │          │                  │          results to response and/or
│ • Prior definitive      │          └──────────────────┘          resistance to treatment
│   therapy (yes vs no)   │                                      • Pharmacokinetics
│ • ICC (platinum/        │                                      • Assessment of
│   fluoropyrimidine vs   │                                        immunogenicity
│   platinum/paclitaxel)  │
└─────────────────────────┘
```

Abbreviations: chemo, chemotherapy; DCR, disease control rate; DOR, duration of response; ESCC, esophageal squamous cell carcinoma; HRQoL, health-related quality of life; ICC, investigator choice of chemotherapy; ORR, objective response rate; OS, overall survival; PFS, progression-free survival; PFS2, PFS after next line of treatment.

After providing written informed consent, completing all screening assessments, and being confirmed as eligible for study participation, eligible patients will be randomized 1:1 to receive either tislelizumab in combination with a chemotherapy doublet (Arm A) or placebo with a chemotherapy doublet (Arm B), by 24 November 2020. The choice of chemotherapy must be determined prior to randomization.

Patient randomization will be stratified by the following 3 factors:
- Geographic region (Asia [excluding Japan] vs Japan vs Rest of World)
- Prior definitive therapy (yes vs no)*
- Investigator choice of chemotherapy (platinum with fluoropyrimidine vs platinum with paclitaxel)

*: Please note that in addition to IRT data, "prior definitive therapy" is also derived using CRF data with the following algorithm: the "treatment intent" of "curative" from 1) prior radiotherapy given in locally advanced setting page or 2) prior surgery related with cancer pages.

After randomization, patients will begin double-blind treatment with one of the following

regimens.

- Arm A:
  - o Tislelizumab + chemotherapy doublet
  - o The chemotherapy doublet will consist of:
    - ▪ Platinum (cisplatin or oxaliplatin) and a fluoropyrimidine (capecitabine or 5-fluorouracil [5-FU])

    OR
    - ▪ Platinum (cisplatin or oxaliplatin) and paclitaxel

- Arm B:
  - o Placebo + chemotherapy doublet
  - o The chemotherapy doublet will consist of:
    - ▪ Platinum (cisplatin or oxaliplatin) and a fluoropyrimidine (capecitabine or 5-FU)

    OR
    - ▪ Platinum (cisplatin or oxaliplatin) and paclitaxel

The platinum agent may be cisplatin or oxaliplatin (except in China (including Taiwan) and Japan where oxaliplatin substitution is not permitted) according to site or investigator preference or standard practice as determined prior to randomization.

Cross-over between treatment arms or switch between fluoropyrimidine and paclitaxel during the study treatment period will not be allowed.

Study treatment will be administered until disease progression, intolerable toxicity, or another reason for treatment discontinuation criterion is met. Platinum therapy may be stopped after 6 cycles, per site or investigator preference or standard practice. If platinum treatment is stopped, the non-platinum agent (fluoropyrimidine or paclitaxel) may continue at the regular schedule.

2.2   STUDY ASSESSMENTS

Tumor response will be assessed by the investigator or by the BIRC for the assessment of the secondary endpoints and exploratory endpoints. Baseline tumor imaging (computed tomography [CT] with or without contrast or magnetic resonance imaging [MRI]) must be performed within 28 days prior to randomization. Post-baseline tumor assessments will occur every 6 weeks (± 7 days) for the first 48 weeks after randomization, then every 9 weeks (± 7 days) based on RECIST v1.1, regardless of treatment delays, until disease progression. Patients who discontinue study treatment early for reasons other than documented radiographic disease progression (eg, toxicity) will continue to undergo tumor assessments following the original plan until the patient experiences radiographic disease progression, withdraws consent, is lost to follow-up, death, or until the study completes, whichever occurs first. Investigators 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.

Patient Reported Outcomes (PROs) of HRQoL will be collected using the EORTC QLQ-C30, EORTC QLQ-OES18, and the EQ-5D-5L at baseline, prior to dosing of every treatment cycle for the first 6 cycles, then every other cycle afterwards, and at the Safety Follow-up Visit.

All AEs will be reported during the study (AEs from the time of the first dose and SAEs from the

VV-CLIN-092227   Version   2.0

time of signing of informed consent) and for up to 30 days after the last dose of study drug(s) (including chemotherapy) or until initiation of another anticancer therapy, whichever occurs first. Immune-mediated AEs (which is equivalent to protocol defined immune-related AEs, serious or non-serious) should be reported up to 90 days after the last dose of tislelizumab, regardless of whether the patient starts a new anticancer therapy. The investigator should report any SAEs that are assessed as related to study treatment, at any time after treatment discontinuation. AEs will be graded according to NCI CTCAE v4.03.

Safety and efficacy monitoring will be performed by an Independent Data Monitoring Committee (IDMC). The IDMC may recommend modifications to the study, including 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 evaluate and compare the overall survival (OS) following treatment with tislelizumab in combination with chemotherapy compared to placebo in combination with chemotherapy when given as first-line treatment in patients with unresectable, locally advanced recurrent or metastatic ESCC

3.2   SECONDARY OBJECTIVES

- To evaluate and compare the efficacy of tislelizumab in combination with chemotherapy compared to placebo in combination with chemotherapy as a first-line treatment in unresectable, locally advanced recurrent or metastatic ESCC as measured by PFS, ORR, and DOR assessed by the investigator per RECIST v1.1

- To evaluate and compare the efficacy of tislelizumab in combination with chemotherapy with the efficacy of placebo in combination with chemotherapy as a first-line treatment in unresectable, locally advanced recurrent or metastatic ESCC as measured by OS in the PD-L1 score ≥ 10% subgroup. PD-L1 expression is determined by PD-L1 score assessed by tumor area positive score (TAP) (previously referred to visually estimated Combined Positive Score [vCPS] in protocol), 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

- To evaluate and compare health-related quality of life (HRQoL) based on patient-reported outcomes (PROs) between tislelizumab in combination with chemotherapy and placebo in combination with chemotherapy

- To compare the safety between tislelizumab in combination with chemotherapy and placebo in combination with chemotherapy

3.3 EXPLORATORY OBJECTIVES

- To characterize the DCR with tislelizumab in combination with chemotherapy assessed by the investigator per RECIST v1.1

- To evaluate PFS, ORR, DOR and DCR assessed by BIRC per RECIST v1.1

- To assess PFS after next line of treatment (PFS2)

- To explore biomarkers in tumor tissues and/or blood samples before and after study treatment and/or at PD/reoccurrence, and the association between these biomarkers and clinical efficacy, disease status, and resistance. Biomarker assessment will consist of PD-L1 expression, gene expression profiling (GEP), tumor mutation burden (TMB)/microsatellite instability (MSI)/mutation profile, and tumor-infiltrating immune cells. Other assessments may be conducted as allowed by local regulations.

- Assessments of pharmacokinetics of tislelizumab when given with chemotherapy

- To assess host immunogenicity to tislelizumab in combination with chemotherapy

4 DEFINITION OF PRIMARY ESTIMAND

The primary scientific question of interest is: will the addition of Tislelizumab to chemotherapy doublet treatment prolong survival in first line ESCC patients, regardless of subsequent anticancer therapy?

The primary estimand is described by the following attributes:

1. Treatment of interest:
   The **treatment of interest** is the randomized treatment (tislelizumab plus chemotherapy doublet as experimental regimen versus placebo plus chemotherapy doublet as control regimen). Doublets include platinum (Cisplatin or Oxaliplatin) with fluoropyrimidine (5-Fluorouracil or Capecitabine) or Paclitaxel.
2. Population:

All subjects randomized in the study--Adult patients with unresectable, locally advanced recurrent or metastatic ESCC, who have not received prior systemic therapy for locally advanced unresectable or metastatic ESCC. Further details on the population characteristics are provided in the protocol eligibility criteria(Details refer to Section 4 in protocol).

3. Primary variable:
   Overall survival (see Section 5.1.1), defined as the time from the date of randomization to the date of death due to any cause
4. Handling of intercurrent events:
   - New anticancer therapy started prior to death: any new anticancer therapy will be ignored (treatment policy strategy)
   - Discontinuation of study treatment: discontinuation of any component of study treatment will be ignored (treatment policy strategy).

- Any unforeseen intercurrent events (e.g., COVID-19 -related events): OS will take into account all deaths irrespective of any unforeseen intercurrent events (treatment policy strategy)

5. <u>Summary measure</u>:

Hazard ratio (HR) of OS comparing tislelizumab plus chemotherapy doublet versus placebo plus chemotherapy doublet, stratified by pooled geographic region* (Asia vs Rest of World), prior definitive therapy (yes vs no) and ICC option (Investigator choice of chemotherapy [platinum with fluoropyrimidine vs platinum with paclitaxel]).

*: Pooled geographic region as "Asia" include "Asia [excluding Japan]" and " Japan" which are defined as geographic regions in study protocol.

5 STUDY ENDPOINTS

5.1.1 Primary Endpoint

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

5.1.2 Secondary Endpoints

- PFS - defined as the time from the date of randomization to the date of first documentation of disease progression assessed by investigator per RECIST v1.1 or death, whichever occurs first

- ORR - defined as the proportion of patients whose best overall response (BOR) is complete response (CR) or partial response (PR) assessed by investigator per RECIST v1.1

- OS in the PD-L1 score $\geq 10\%$ subgroup

- DOR- defined as the time from the first determination of an objective response until the first documentation of progression assessed by investigator per RECIST v1.1 or death, whichever comes first

- HRQoL assessment of the patient's overall health status using European EORTC QLQ-C30 index, the European esophageal cancer specific module EORTC QLQ-OES18, and the generic health state instrument EuroQol 5D (EQ-5D-5L)

- The incidence and severity of treatment-emergent adverse events (TEAEs) according to National Cancer Institute Common Terminology Criteria for Adverse Events (NCI-CTCAE) v4.03

5.1.3 Exploratory Endpoints

- DCR - defined as the proportion of patients whose BOR is CR, PR, and stable disease (SD) assessed by investigator per RECIST v1.1

- PFS, ORR, DOR and DCR assessed by the BIRC per RECIST v1.1

- PFS2 - defined as the time from randomization to the objective disease progression after next line of treatment or death from any cause, whichever occurs first.

- To explore biomarkers including but not limited to PD-L1 expression, gene expression profiling (GEP), tumor-infiltrating lymphocytes (TIL), tumor mutation profile/tumor mutation burden (TMB)/microsatellite instability-high (MSI-H) status in tumor tissue and/or blood sample before, after study treatment, and/or at progressive disease (PD)/reoccurrence, and the association between these biomarkers and clinical efficacy, disease status, and resistance

- Assessments of pharmacokinetics of tislelizumab when given with chemotherapy

- Assessments of immunogenicity of tislelizumab to determine the incidence of anti-drug antibodies (ADA)

6 SAMPLE SIZE CONSIDERATIONS

The initial sample size calculation was based on the primary efficacy analysis of PFS and OS in the comparison between tislelizumab in combination with chemotherapy arm and placebo plus chemotherapy arm in the ITT analysis set. The Type I error was strongly controlled by initially assigning a one-sided alpha of 0.005 to the PFS hypothesis and 0.02 to the OS hypothesis. Hazard ratios in PFS and OS were assumed as 0.65 and 0.73, respectively, with median PFS of 5 months and OS of 9 months in the comparator arm. A total of 480 patients would be enrolled in a 1:1 randomization over a 17-month period at enrollment rates of 10 patients/month in the first month, 20 patients/month in the second month and 30 patients/month in the next 15 months. Approximately 319 PFS events were planned in the PFS hypothesis testing to have a power of 90% with an one sided alpha of 0.005. A group sequential testing of OS would be performed. The interim analysis was planned after approximately 67% of the total planned death events had occurred (241). The final analysis of OS would be performed when approximately 360 death events have been observed. The sample size calculation of OS was based on overall power of 82% and an one-sided alpha level of 0.02.

The protocol amendment 3.0 was amended on 25 May 2020. The PFS and OS hazard ratio assumptions was updated after reviewing the recently published results of anti-PD-1 therapies in the second line treatment of ESCC (Kojima et al 2019, Kato et al 2019, Huang et al 2019) which showed delayed treatment effect and increased usage of subsequent immunotherapies. Using a more conservative OS HR assumption of 0.76 around the time of final analysis after an initial 3-month delayed treatment effect (e.g. assuming HR=1 in the first 3 months), the number of deaths required in the final analysis is increased from approximately 360 to 467 events for an 81% power using simulation. The updated number of deaths planned in the interim analysis with the aforementioned statistical methodology is increased from approximately 241 to 313 events. Similarly, with a PFS HR assumption of 0.68 around the time of interim analysis after an initial 3-month delayed treatment effect, the PFS events are increased from approximately 319 to 423 events to achieve an estimated power of 92% in the interim analysis (final analysis for PFS) using simulation. The interim analysis will be carried out after both targeted PFS and death events are

met. The sample size is increased to 622 patients who will be enrolled over an estimated 26-month period. A 5% annual dropout rate is assumed in the sample size calculation.

By Nov 24, 2020, enrollment was completed with 649 randomized patients. In the current protocol amendment 4.0, PFS per BIRC is dropped from the primary efficacy analysis as it is no longer considered as approvable endpoint given the positive OS results from KEYNOTE-590 (Kato et al 2020) in the 1st line treatment of ESCC. OS is kept as the sole primary efficacy endpoint. OS HR is assumed to be 0.74 at the time of final analysis after an initial 1-month delayed treatment effect (e.g. assuming HR=1 in the first month), the number of deaths required in the final analysis will be approximately 488 with 90% power by simulation. The planned interim analysis with a power of 81% will occur after 423 events have been observed. A 5% annual dropout rate is assumed and one-sided alpha of 0.025 is used in the sample size calculation.

7 STATISTICAL METHODS

7.1 ANALYSIS POPULATIONS

- The Intention-to-Treat (ITT) analysis set includes all randomized patients. It will be the primary analysis population for the efficacy analysis

- The Safety analysis set includes all patients who received at least 1 dose of study drug. It will be the primary analysis population for safety analysis.

- The Pharmacokinetics (PK) analysis set includes all patients who are receive at least 1 dose of tislelizumab per the protocol, for whom any post-dose PK data are available

- The ADA analysis set includes all patients who have a baseline and at least 1 post-baseline ADA result

7.2 DATA ANALYSIS GENERAL CONSIDERATIONS

7.2.1 Definitions and Computations

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

In the situation where the event date is partial or missing, the date will appear partial or missing in the listings; Study day and any corresponding durations will be presented based on the imputations specified in Appendix 11.

Baseline Measurements:

- Baseline characteristics and efficacy analysis including HRQol: a baseline value is defined as the last non-missing value collected up to the randomization.

- Safety analysis: a baseline value is defined as the last non-missing value up to the first study drug administration.

Study Follow-up Duration (SFD): Study follow-up duration is defined as the duration from the randomization date to death date if the patient died prior to data cutoff, or the study discontinuation date if the patient discontinued from study prior to data cutoff, otherwise to the data cutoff date.

Minimum study follow up (MinFU) is defined as a difference between the date of cutoff and the date of last patient randomized.

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

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.

- 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 will be based on the actual date 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, Q1, Q3 and range (minimum and maximum).

7.2.3 Handling of Missing Data

Missing data will not be imputed unless otherwise specified elsewhere in the SAP. This included withdraw from study by subjects which have no further data collection. 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 11.1, 11.2, 11.3 and 11.4

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.

7.2.4 Adjustments for Covariates / Stratification

The stratification factors used for randomization (IRT recorded), including pooled geographic region (Asia vs Rest of World), prior definitive therapy (yes vs no) and ICC option (Investigator choice of chemotherapy [platinum with fluoropyrimidine vs platinum with paclitaxel]), will be used in stratified log-rank test and stratified Cox proportional hazard model for primary endpoint OS and secondary endpoints PFS and other secondary endpoints. Similarly these stratification factors will be used in Cochran-Mantel-Haenszel method to analyze ORR.

7.2.5 Multiplicity Adjustment

The type I error will be strongly controlled at 0.025 (1-sided) in the primary analysis of OS in the ITT analysis set.

There will be 1 interim analysis of OS utilizing the O'Brien-Fleming boundary approximated by Hwang-Shih-DeCani spending function with the gamma parameter set at -4. Details may refer to Section 8.

By using the graphical approach of Bretz et al 2009, if the null hypothesis for OS in the ITT analysis set is rejected, the corresponding alpha will be shifted to the hypothesis tests of the secondary endpoints, PFS by investigator in ITT, ORR by investigator in ITT, OS in PD-L1 score$\geq$10% subgroup and HRQoL in ITT, which will be tested sequentially. The inferential test will be stopped at the first non-significant endpoint. For HRQoL, Bonferroni method will be applied to the test of OES 3 symptoms dysphagia, eating and reflux using an alpha of 0.0083. To control overall type I error in the secondary endpoint testing, all secondary endpoints will be tested only once using data from the interim analysis with an alpha of 0.025 (0.0083 for each of the beforementioned OES symptoms), no matter when the null hypothesis of OS in the primary analysis is rejected (IA or final). Therefore, if OS superiority is demonstrated only at the final analysis, the results of secondary endpoints from the interim analysis will be used for hypothesis testing, while those data from the final analysis will be summarized for descriptive purpose only.

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.

7.3 SUBJECT CHARACTERISTICS

7.3.1 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 vs Rest of World) for ITT analysis set.

7.3.2 Protocol Deviations

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

7.3.3 Demographic and Other Baseline Characteristics

Demographic and other baseline characteristics will be summarized in the ITT Population using descriptive statistics.

Demographic and other baseline characteristics include:

- Age
- Age group (<65 vs >= 65 years)
- Sex
- Race
- Ethnicity
- Height (cm)
- Weight (kg)
- BMI (kg/m^2)
- Region: (Asia: Japan, ex-Japan); Rest of World
- Tobacco use status
- Alcohol consumption status
- ECOG status

.

The demographic summary will also be repeated by region (Asia and Rest of World) and by

baseline PD-L1 status in ITT population.

In addition, the stratification factors per IRT and per eCRF will be summarized based on ITT population:

- Prior definitive therapy (yes vs no)
- Geography (Asia [excluding Japan] vs Japan vs Rest of World)
- Choice of chemotherapy doublet (platinum plus fluoropyrimidine vs platinum plus paclitaxel)

7.3.4 Disease History and Baseline Disease Characteristics

Following disease history and characteristics at study entry will be summarized using ITT analysis set:

- Time since initial cancer diagnosis to Study Entry (month)
- Time since metastatic disease diagnosis to date of Study Entry (month)
- Primary site of esophageal cancer
- Histological Tumor differentiation grade and type
- Disease stage at diagnosis
- Metastatic disease status at study entry
- Metastatic location at study entry
- Number of metastatic sites involved (0, 1, 2, ≥ 3) at study entry
- Target lesions sum of diameters by investigator
- PD-L1 status (PD-L1 score $\geq 10\%$, $< 10\%$, Unknown)

Histology Grade Cancer associated symptoms at baseline will be coded by MedDRA version 24.0 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 Rest of World) and baseline PD-L1 status in ITT analysis set

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

7.3.5 Prior Anti-Cancer Therapies and Surgeries

The number of subjects receiving prior anti-cancer drug therapies, prior anti-cancer radiation therapy including prior chemo-radiation therapy, prior anti-cancer surgeries will be summarized. The therapies with the same sequence/regimen number are counted as one prior therapy. The number of subjects with at least one prior definitive therapy, and its subcategories, definitive surgery with/without adjuvant/neo-adjuvant treatment and definitive radiotherapy with/without chemotherapy, will also be summarized. In addition, time from end of last anti-cancer therapy to study entry will be summarized.

The prior anti-cancer therapy will also be summarized by region (Asia and Rest of World) and by baseline PD-L1 status for ITT analysis set.

7.3.5.1   Prior Anti-Cancer Systemic Therapy

The number (percentage) of patient with at least one prior anti-cancer systemic therapy, platinum based prior systemic therapy, treatment setting of prior anti-cancer systemic therapies, and duration of last prior anti-cancer systemic therapy (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).

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

7.3.6   Prior and Concomitant Medication and Therapy

Prior and concomitant medications will be coded using the World Health Organization Drug Dictionary (WHO DD) drug codes, and will be further classified to the appropriate Anatomical Therapeutic Chemical (ATC) code.

The number (percentage) of subjects reporting prior and concomitant medications will be summarized by ATC medication class and WHO DD preferred term by phase in the safety population. Prior medications are defined as medications that received within 30 days before randomization and stopped before the first dose date. Concomitant medications will be 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 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.

Medication to treat COVID-19-related adverse events will also be summarized separately. A listing of prior and concomitant medications will be provided

7.3.7   Medical History

Medical History will be coded using MedDRA (version 24.0 or newer). 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 in the safety population. A listing of medical history will be provided.

7.4   EFFICACY ANALYSIS

7.4.1   Primary Efficacy Endpoint

7.4.1.1   Primary Estimand

Primary estimand is defined in Section 4.

7.4.1.2   Primary analysis for primary estimand

Overall survival is defined as time from randomization date to the documented death date for

patients who died prior to or on the clinical cutoff date. For patients who are alive by the clinical cutoff date, OS will be censored at the last known alive date (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 cutoff 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. Any deaths after missing visits are still considered as events.

<u>Overall Survival in ITT</u>

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

$$H_0 : HR \geq 1,$$

$$H_a : HR < 1,$$

where HR represents the hazard ratio between the treatment and the control groups, assuming a proportional hazard. The null hypothesis will be tested using a log-rank test stratified by pooled geographic region (Asia vs Rest of World), prior definitive therapy (yes vs no) and ICC option (Investigator choice of chemotherapy). If the one-sided p-value is less than significance level threshold specified in table 7 for interim or final analysis, it will be concluded that the null hypothesis is rejected and the superiority of treatment group over the control group in OS is demonstrated at the significance level.

The HR and its 2-sided 95% CI will be estimated from a stratified Cox regression model with the same stratification factors above. The stratified hazard ratio will be the main estimator for the primary estimand of OS. The exact method will be used for handling the tied events. The distribution of OS, including median, Q1 and Q3, and event-free rates at *every 3 months* will be estimated using the Kaplan-Meier method for each treatment group. Ninety-five percent CIs for median and Q1 and Q3 of OS will be estimated using the method of Brookmeyer and Crowley (Brookmeyer and Crowley, 1982). And 95% CIs for event-free rates will be estimated using Greenwood's formula (Kalbfleisch and Prentice 1980). Kaplan-Meier survival curve for each arm will also be provided.

OS Sensitivity analyses and supportive analysis:
➢ Analysis 1 "Unstratified OS sensitivity analysis": To assess the impact of stratification, the unstratified HR together with the associated 95% confidence interval obtained using the unstratified Cox regression model will be presented.

➢ Analysis 2 "OS sensitivity analysis with stratification factors from the eCRF": To assess the impact of mis-stratification, the HR together with the associated 95% confidence interval obtained using the stratified Cox regression model with stratification factors collected in the eCRF will be presented.

➢ Analysis 3 "OS Supportive analysis of Proportional hazard assumption": Analyses to assess proportional hazard assumption including Schoenfeld residual plot and time dependent covariate using treatment by time interaction in the cox model will be explored.

Supplementary analyses related to OS

Supplementary analysis 1 "OS analysis based on Restricted mean survival time method": This analysis targets an estimand which has the same attributes as the primary estimand except for the summary measure will be difference in Restricted mean survival time (RMST) between two treatment groups. In order to account for the possible non-proportional hazard effect, the restricted mean survival time (RMST) (RMST, Uno H, Claggett B, Tian L, Inoue E, et al. 2014) will be computed for OS 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 2 "OS analysis based on Max-Combo method": This analysis targets an estimand which has the same attributes as the primary estimand except for the summary measure which will be hazard ratio from combination of Fleming and Harrington weighted log-rank test (FH) test (Max-Combo) based on the $G^{\rho,\gamma}$ family between two treatment groups. In order to account for the possible non-proportional hazard effect, combination of $G^{0,0}$, $G^{0,1}$, $G^{1,1}$, $G^{1,0}$ will be computed which allows possibility for different NPH type.

Supplementary analysis 3 'OS analysis adjusted for baseline covariates': The target population, the primary variable, the treatment of interest, intercurrent events are the same as for the primary estimand. A Cox regression model stratified by randomization stratification factors will be fitted to evaluate the effect of other baseline demographic and disease characteristics on the estimated hazard ratio. The fitted model adjusting the treatment difference for key baseline and prognostic factors will include as covariates the following: baseline PD-L1 status (PD-L1 score $\geq$ 10%, PD-L1 score < 10%), age(<65, >=65), sex, race, smoking, ECOG status, metastatic (yes or no)

Supplementary analysis 4 "COVID-19 OS supplementary analysis": This analysis may be performed when there are sufficient patients (i.e., 5% of ITT) meeting the criteria below regarding COVID-19. This analysis will aim at assessing the treatment effect based on OS had COVID-19 pandemic not occurred will also be conducted. The target population, treatment of interest and the summary measure of this endpoint are the same as for the primary estimand. The primary variable is defined as the time from the date of randomization to the date of death due to non-COVID-19 pandemic reasons. The remaining intercurrent events will be handled as follows:

- **Discontinuation of study treatment due to any non-COVID-19 pandemic reasons**: OS will take into account all deaths irrespective of the study treatment discontinuation reasons (treatment policy)
- **Discontinuation of study treatment due to COVID-19 pandemic reasons**: OS will be censored on the date of discontinuation of treatment due to COVID-19 pandemic (hypothetical strategy). The discontinuation reason due to COVID-19 pandemic will be identified from the defined COVID-19 protocol deviations.
- **Medications used for treating COVID-19 cases:** OS will be censored on the date of administration of COVID-19 medication (hypothetical strategy). The medications will be identified by drug class of aminoquinolines, protease inhibitors, glucocorticoids,

macrolides, antiviral or antiretroviral OR drug names that contain key word of "hydroxychloroquine" or "chloroquine".

o **Death due to COVID-19**: OS will be censored on the date of death due to COVID-19 (hypothetical strategy)

VV-CLIN-092227   Version 2.0

Supplementary analyses targeting different estimands are described in Table 1.

Table 1 Supplementary analyses for OS

	Primary estimand	Supplementary estimand 1	Supplementary estimand 2	Supplementary estimand 3	Supplementary estimand 4
Scientific question of interest	Will the addition of Tislelizumab to chemotherapy doublet treatment strategy prolong survival in first line ESCC patients	same as primary estimand	same as primary estimand	same as primary estimand	Will the addition of Tislelizumab to chemotherapy doublet treatment strategy prolong survival due to non-COVID-19 pandemic reasons in first line ESCC patients
Treatment	Arm A and Arm B	same as primary estimand	same as primary estimand	same as primary estimand	same as primary estimand
Population	Unresectable, Locally Advanced Recurrent or Metastatic Esophageal	same as primary estimand	same as primary estimand	same as primary estimand	same as primary estimand

	Squamous Cell Carcinoma				
Intercurrent events					
- New anticancer therapy	Ignored (treatment policy strategy)	same as primary estimand	same as primary estimand	same as primary estimand	same as primary estimand
- Discontinuation of treatment;	Ignored (treatment policy strategy)	same as primary estimand	same as primary estimand	same as primary estimand	same as primary estimand
Summary measure	Hazard ratio of OS from stratified cox model	Difference in Restricted mean survival time	Weighted hazard ratio of OS from Max-combo test	Hazard ratio of OS from stratified cox model adjusted by covariates	same as primary estimand
Methods used for analysis	Stratified Cox regression analysis	RMST	Max-combo	Multivariate analysis	Stratified Cox regression analysis

7.4.2 Secondary Efficacy Endpoints

Progression Free Survival (PFS) by investigator in ITT

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

$$H_0 : HR \geq 1,$$

$$H_a : HR < 1,$$

where HR represents the hazard ratio between the treatment and the control groups, assuming a proportional hazard. After superiority of OS in the ITT analysis set has been demonstrated, the null hypothesis will be tested using a log-rank test stratified by pooled Geographic region (Asia vs Rest of World), prior definitive therapy (yes vs no) and ICC option (Investigator choice of chemotherapy [platinum with fluoropyrimidine vs platinum with paclitaxel]). If the one-sided p-value is less than *0.025*, it will be concluded that the null hypothesis is rejected and the superiority of treatment group over the control group in PFS is demonstrated at the significance level.

The HR and its 2-sided 95% CI will be estimated from a stratified Cox regression model with the same stratification factors above. The distribution of PFS, including median, Q1 and Q3, and event-free rates at every 3 months, will be estimated using the Kaplan-Meier method for each treatment group. Ninety-five percent CIs for median and Q1 and Q3 of PFS will be estimated using the method of Brookmeyer and Crowley (Brookmeyer and Crowley, 1982). And 95% CIs for event-free rates will be estimated using Greenwood's formula (Kalbfleisch and Prentice 1980). Kaplan-Meier survival curve for each arm will also be provided.

Progression-Free survival is 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. Progression-free survival censoring rule for PFS primary and supportive analysis are described in Table 2.

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

No.	Situation	Date of Progression or Censoring	Primary Analysis	Supportive Analysis 1	Supportive Analysis 2
1	No baseline or any post-baseline tumor assessments and without death within 13 weeks after randomization	Randomization date	Censored	Censored	Censored
2	Progression documented	Date of first radiologic PD assessment	Event	Event	Event***

	between scheduled visits				
3	No progression at the time of data cutoff or withdrawal from study	Date of last adequate radiologic assessment prior to or on date of data cutoff 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	NA	Censored
5	Death before first PD assessment	Date of death	Event	Event	Event***
6	Death between adequate assessment visits*	Date of death	Event	Event	Event***
7	Death or progression after more than one missed visit**	Date of last adequate radiologic assessment before missed tumor assessments	Censored	Censored	Event
8	No baseline or any post-baseline tumor assessments and died within 13 weeks after randomization	Date of death	Event	Event	Event***

*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 48 weeks and once every 9 weeks afterwards with one-week window, D2 is 12 weeks + 1 week in the first 48 weeks 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. Note 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 supportive analysis 1 in table 2, the PFS event date will be derived ignoring new anti-cancer therapy.

4. In supportive analysis 2 in table 2, any PD or death after more than one missing tumor assessment will be considered as a PFS event.

<u>Objective Response Rate (ORR) by investigator in ITT</u>

The null and alternative hypotheses for ORR are set as follows:

$$H_0 : ORR_T \leq ORR_{C'}$$

$$H_a : ORR_T > ORR_{C'}$$

where ORR_T and ORR_C represent ORR of the Tislelizumab treatment and Control groups, respectively. The null hypothesis will be tested using the Cochran-Mantel-Haenszel method stratified by pooled Geographic region (Asia vs Rest of World), prior definitive therapy (yes vs no) and ICC option (Investigator choice of chemotherapy [platinum with fluoropyrimidine vs platinum with paclitaxel]) at the one-sided significance level of 0.025. If the null hypothesis can be rejected, it will be concluded that the superiority of treatment group over the control group in ORR is demonstrated at the significance level.

Patients with no post-baseline response assessment (for any reason) will be considered non-responders. The 2-sided 95% CIs for the odds ratio in ORR will be calculated, as well as Clopper-Pearson 95% CIs of ORR for each treatment arm.

Best overall response, defined as the best response recorded from randomization until data cut 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. The proportion and its corresponding Clopper-Pearson 95% CI for each of the response categories (CR, PR, SD, and PD) will be presented by treatment arm.

Overall Survival in ITT PD-L1 (PD-L1 score ≥ 10%) subgroup

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

$$H_0 : HR \geq 1,$$

$$H_a : HR < 1,$$

where HR represents the hazard ratio between the treatment and the control groups in ITT PD-L1 score ≥ 10% subgroup, assuming a proportional hazard.

The PD-L1 score ≥ 10% subgroup includes patients whose baseline PD-L1score ≥ 10% using VENTAN A PD- L1 (SP263) CDx Assay.

The overall survival analysis by other baseline PD-L1 status (PD-L1 score < 10 %) are analyzed similarly as in ITT analysis set.

Duration of Response (DOR) assessed by investigator in ITT
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 2 Table). should be applied to DOR as well.

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.

7.4.3 Subsequent Anti-Cancer Therapy

To assess the subsequent anti-cancer therapy within two treatment groups which potentially impact patient overall survival, 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 by baseline PD-L1 status (PD-L1 score ≥ 10%, <10%, Unknown).

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 could be the earlist of date of the new 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 deem appropriate.

- The start date of new anti-cancer therapy in defining TEAE for safety analyses is always the first date of new systemic anti-cancer therapy taken after the last study treatment.

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

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

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

7.4.4 Exploratory Analysis

Disease control rate (DCR) assessed by investigator.

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 investigators will be analyzed similarly to ORR in the ITT

PFS, ORR, DOR assessed by BIRC in ITT

PFS, ORR, DOR, and DCR assessed by BIRC per RECIST v1.1 will be summarized using the same approach as by investigator in the exploratory analysis. No formal testing will be performed for BIRC assessed endpoints.

The concordance/discordance of PFS by BIRC and investigators will be assessed by comparing PFS event type/censor between BIRC and investigators, as well as summary of gap time in months for PFS as per BIRC and as per investigators.

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 (ND) at subsequent tumor assessment, provided the criteria for PD or NE are not met. In that case, ND or PD will be considered as valid tumor assessment when performing the analysis of PFS.

PFS2 assessed by investigator in ITT

Analysis of progression-free survival after next line of treatment (PFS2) is defined as the time from randomization to second/subsequent disease progression, or death from any cause, whichever occurs first. To calculate PFS2, data from patients without disease progression after next-line of treatment or death at the time of analysis will be censored at the last time known to be alive. Kaplan-Meier (KM) method as described in the PFS analyses will be used in the analysis of PFS2.

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

study treatment, or patient has discontinued study treatment but has not started next-line therapy and is still alive.

- Any death prior to initiation of next-line therapy will be considered as an event for PFS2

7.4.5 Health-Related Quality of Life

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) ÷ # non-missing of DS

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] ÷ # non-missing of DS

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

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

For EQ-5D-5L, descriptive modules 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.

In addition, a mixed effect model analysis for measuring clinically meaningful changes post-baseline adjusted for baseline will be performed using the GHS, physical function and fatigue domains of QLQ-C30 and index score, dysphagia, reflux, pain and eating of QLQ-OES18. Treatment effect of change from baseline will be evaluated at cycle 6. Bonferrroni method will be applied to the test of OES 18 symptoms dysphagia, eating and reflux using an alpha of 0.0083.

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. And a secondary deterioration threshold based on data may be explored.

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 pooled Geographic region (Asia vs Rest of World), prior definitive therapy (yes vs no) and ICC option (Investigator choice of chemotherapy [platinum with fluoropyrimidine vs platinum with paclitaxel]).

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

7.4.6 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. Rest of World
  - Asia (excluding Japan) vs. Japan vs. Rest of World
- ECOG Performance status (0/1)
- Age group (<65 years, >= 65 years)
- Sex
- Smoking status (former/current smoker, non-smoker)

- Race (White, Asian and Other)

- Locally advanced vs. metastatic

- Prior definitive therapy (yes/no)

- Baseline PD-L1 expression category: PD-L1 score>= 10%, PD-L1 score< 10%, Unknown

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, DOR, HRQoL and subsequent anti-cancer therapy will be summarized by region (Asia vs Rest of World) in ITT analysis set.

PFS, ORR, DOR assessed by investigator will be summarized by baseline PD-L1 status (PD-L1 score ≥ 10%, < 10 %, Unknown) using the same approach in ITT analysis set.

7.5 SAFETY ANALYSES

Safety will be assessed by monitoring and recording of all TEAEs graded by NCI-CTCAE v4.03. Laboratory values (e.g., hematology, clinical chemistry), vital signs, ECGs, PEs and their changes from baseline will also be used in determining 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.

Extent of Exposure

Extent of exposure will be summarized by treatment arm in ITT.

For control 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 of patients (if any) treated beyond radiological progression (as per investigator) will also be presented.

The number (percentage) of patients requiring infusion related infusion interruption/infusion rate decreased, dose delay, and dose reduction (in the Chemotherapy doublets 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 the Chemotherapy doublets.

- Actual dose intensity is defined as 21*total cumulative dose (mg) / (last dose date up to cutoff date+ 21 - first dose date). Relative dose intensity is defined as Actual dose intensity

/ planned dose intensity. The planned dose intensity is the total planned dose in a cycle. For a patient in the tislelizumab/Placebo arm, the planned dose intensity is 200 mg Q3W.

- In case there are some patients received the study drug different from assigned treatment group in safety analysis set, dose summary will only include the data as planned.

The derivations of ADI, planned dose and RDI for Chemotherapy are shown in Table5.

Table 5 ADI, planned dose and RDI for Chemotherapy

	ADI(mg/m²/cycle)	PDI(Planned dose per cycle)	RDI
Paclitaxel	$$\dfrac{\sum_1^{\#of\ cycles} \dfrac{actual\ dose\ (mg)}{BSA*(m2)} \times 21}{date\ of\ last\ dose\ up\ to\ cutoff + 21 - first\ dose\ date}$$	First non-missing planned dose level collected in CRF	$\dfrac{ADI}{PDI}$
Cisplatin**	$$\dfrac{\sum_1^{\#of\ cycles} \dfrac{actual\ dose\ (mg)}{BSA*(m2)}}{max(\dfrac{date\ of\ last\ dose\ up\ to\ cutoff + 21\ or\ 19 - first\ dose\ date}{21}, number\ of\ cycles\ in\ last\ dosing\ CRF\ page)}$$	First non-missing planned dose level collected in CRF	
Oxaliplatin	$$\dfrac{\sum_1^{\#of\ cycles} \dfrac{actual\ dose\ (mg)}{BSA*(m2)} \times 21}{date\ of\ last\ dose\ up\ to\ cutoff + 21 - first\ dose\ date}$$	First non-missing planned dose level collected in CRF	
5-FU	$$\dfrac{\sum_1^{\#of\ cycles} \dfrac{actual\ dose\ (mg)}{BSA*(m2)}}{max(\dfrac{date\ of\ last\ dose\ up\ to\ cutoff + 17 - first\ dose\ date}{21}, number\ of\ cycles\ in\ last\ dosing\ CRF\ page)}$$	First non-missing planned does level collected in CRF*5	
capecitabine	$$\dfrac{\sum_1^{\#of\ cycles} \dfrac{actual\ dose\ (mg)}{BSA*(m2)}}{max(\dfrac{date\ of\ last\ dose\ up\ to\ cutoff + 8 - first\ dose\ date}{21}, number\ of\ cycles\ in\ last\ dosing\ CRF\ page)}$$	First non-missing planned does level collected in CRF*14	$\dfrac{ADI}{28000}$

*BSA: sqrt(height(cm)*weight (kg)/3600) or sqrt(height(cm)*weight (lb)/3131)

** Cisplatin: apply 21 if value of CRF field "Cisplatin administrated days number" value is "1 day" and 19 if value is "3 days".

Treatment duration

If patients discontinued treatment (with non-missing EOT date):

- For the tislelizumab arm, Treatment duration (days) = minimum of (date of the last dose + 20, death date, cutoff date) – date of first dose + 1.

- For patients taking cisplatin or oxaliplatin or Paclitaxel on a Q3W schedule, Treatment duration = minimum of (date of the last dose +20, death date, cutoff date) – date of first dose + 1.

    - For Chemotherapy Doublet C, cisplatin may be administered on Days 1 or 2, or in 3 divided doses on Days 1, 2, and 3 depending on local guidelines.

    - Treatment duration = minimum of (date of the last dose + 20 (for Day 1 or Day 2 schedule) or 18 (for 3 days schedule), death date, cutoff date)– date of first dose + 1respectively.

- For patients taking 5-Fluorouracil, Treatment duration = minimum of (date of the last dose + 16, death date, cutoff date) –date of first dose + 1.

- For patients taking capecitabine, Treatment duration = minimum of (date of the last dose +7, death date, cutoff date) – date of first dose + 1.

Otherwise if patient has treatment ongoing: Treatment duration = Cutoff date– date of first dose + 1

7.5.1 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 24.0 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.

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 study drug up to 30 days following study drug 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 AEs, treatment emergent or otherwise, will be presented in patient data listings. COVID-19 related adverse events will be summarized separately.

It is noteworthy that the definitions of TEAE in the protocol is different from the one in this SAP, while the definition of imAE remains the same. The update of TEAE window streamlines the TEAE derivation so all TEAEs can be identified programmatically instead of relying on the

manual medically review of imAE. imAE occurs outside of the above mentioned TEAE window will not be classified as treatment-emergent adverse events. All imAE will be reported separately.

7.5.1.1 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 TEAE related to any component of study drug

- At least one TEAE related to Tislelizumab

- At least one TEAE related to Tislelizumab and grade ≥3
- At least one TEAE with grade ≥3 and related to any component of study drug
- At least one serious TEAE

- At least one serious TEAE related to tislelizumab

- At least one serious TEAE related to any component of study drug
- At least one TEAE leading to death including death due to disease under study

- At least one TEAE leading to death

- At least one TEAE leading to permanent discontinuation of any component of study drug

- At least one TEAE leading to dose modification of any component 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 PT
  - All TEAEs by SOC and PT*
  - Most frequently reported (incidence ≥ 10% in any treatment arm) TEAEs by SOC and PT
  - Treatment-related TEAEs by SOC
  - Treatment-related TEAEs by PT
  - Treatment-related TEAEs by SOC and PT*

- o Most frequently reported (incidence ≥ 5% in any treatment arm) Treatment-related TEAE by SOC and PT
  - o Tislelizumab/Placebo Related TEAEs by PT
  - o Tislelizumab/Placebo Related TEAEs by SOC and PT*
  - o Most frequently reported (incidence ≥ 10% in any treatment arm) Tislelizumab/Placebo Related TEAE by SOC and PT
  - o Any Chemotherapy Component Related TEAEs by PT
  - o Any Chemotherapy Component Related TEAEs by SOC and PT*
- Serious TEAEs by SOC and PT*
  - o Serious TEAEs by PT
  - o Most frequently reported (incidence ≥ 2% in any treatment arm) serious TEAE by SOC and PT
  - o Treatment-related Serious TEAE by SOC and PT*
  - o Tislelizumab/Placebo Related Serious TEAE by SOC and PT*
  - o Most frequently reported (incidence ≥ 2% in any treatment arm) Tislelizumab/Placebo -related serious TEAE by SOC and PT
  - o Any Chemotherapy Component Related Serious TEAE by SOC and PT*
- TEAEs with NCI-CTCAE grade ≥3 by SOC and PT*
  - o Most frequently reported (incidence ≥ 5% in any treatment arm) TEAE with NCI-CTCAE grade ≥ 3 by SOC and PT
  - o Treatment-related TEAE with NCI-CTCAE grade ≥ 3 by SOC and PT*
  - o Tislelizumab/Placebo Related TEAE with NCI-CTCAE grade ≥ 3 by SOC and PT*
  - o Most frequently reported (incidence ≥ 1% in any treatment arm) Tislelizumab/Placebo Related TEAE with NCI-CTCAE grade ≥ 3 by SOC and PT
  - o Any Chemotherapy Component Related TEAE with NCI-CTCAE grade ≥ 3 by SOC and PT*
  - o Tislelizumab/Placebo Related TEAEs leading to death by SOC and PT*
  - o Tislelizumab/Placebo Related TEAEs leading to death by SOC and PT including death due to disease under study*
  - o Any Chemotherapy Component Related TEAE Leading to Death by SOC and PT*
  - o Any Chemotherapy Component Related TEAE Leading to Death by SOC and PT including death due to disease under study
- TEAEs leading to treatment discontinuation by SOC and PT*

- o TEAE Leading to Treatment Discontinuation of Tislelizumab/Placebo by SOC and PT*

- o Most frequently reported (incidence ≥ 2 patients in any treatment arm) TEAE Leading to Treatment Discontinuation of Tislelizumab/Placebo by PT

- o Treatment-related TEAE Leading to Treatment Discontinuation of Any Chemotherapy Component by SOC and PT*

- o Most frequently reported (incidence ≥ 2 patients in any treatment arm) TEAE Leading to Treatment Discontinuation of Any Chemotherapy Component by PT

- o TEAE Leading to Treatment Discontinuation of All Treatment Components by SOC and PT*

- o TEAE Leading to Treatment Discontinuation of Tislelizumab/Placebo Excluding Treatment Discontinuation Due to Progressive Disease Progression by SOC and PT*

- • TEAEs leading to dose modification by SOC and PT

  - o Treatment-related TEAE Leading to Dose Modification by SOC and PT

  - o TEAE Leading to Dose Modification of Tislelizumab/Placebo by SOC and PT*

  - o Most frequently reported (incidence ≥ 2% in any treatment arm) TEAE Leading to Dose Modification of Tislelizumab/Placebo by SOC and PT

  - o TEAE Leading to Dose Modification of Any Chemotherapy Component by SOC and PT*

  - o Most frequently reported (incidence ≥ 2% in any treatment arm) TEAE Leading to Dose Modification of Any Chemotherapy Component by SOC and PT

The types of Dose modification include dose delayed, infusion interruption, infusion rate decreased and dose reduction for chemotherapy doublets; dose delayed and infusion interruption, infusion rate decreased for tislelizumab arm.

*: Summaries will be provided by region (Asia vs. Rest of World) in addition to safety analysis set.

For exposure adjusted event rate (EAER), overall summary (at least one (serious) TEAE, (serious) treatment-related TEAE, grade ≥3 TEAE, grade ≥3 treatment-related TEAE, TEAE leading to death, treatment-related TEAE leading to death); summary by System Organ Class and Preferred Term will be provided for safety analysis set and by Region.

EAER was calculated as event count*100 / person-months of exposure (longest treatment duration of all study medication components in treatment regimen defined by treatment duration in Section 7.5).

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

7.5.1.2 Immune-mediated Adverse Event

Immune-mediated adverse events are of special interest and summarized by category within a pre-defined list in Appendix 11.5.

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/placebo column.

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

- Immune-mediated adverse events by category and by PT
- Immune-mediated adverse events by category and by PT by region
- Immune-mediated adverse events with NCI-CTCAE grade ≥3 by category
- 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

Exposure-adjusted event rate (EAER) per 100 subject years are provided. The EAER per 100 subject -years/months is defined as 100 times the number of specific events divided by the total exposure time (in years/months) among subjects included in the analysis. Subjects with multiple occurrences of the specific event in the specific analysis period will be counted multiple times in the numerator. The exposure time for a subject with/without the specific event is the treatment duration. The total exposure time in years/months is calculated by dividing the sum of exposure time in days over all subjects included in the analysis by 365.25/30.4375. The EAER per 100 subject-years/months is interpreted as expected number of specific event per 100 subject-years/months of exposure to the study dru g.

7.5.1.3 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/placebo column. Summaries of IRRs, IRRs with NCI-CTCAE grade ≥3, IRRs of tislelizumab/placebo leading to treatment discontinuation, and IRRs of tislelizumab/placebo leading to dose modification will also be provided by PT only, in descending order.

7.5.2 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 safety analysis set

Patient data listings of death and reason will be provided.

7.5.3 Laboratory Values

Hematology, serum chemistry, thyroid function, will be summarized/listed for selected

parameters described in Table 6. The coagulation, urinalysis, HBV/HCV serology, and pregnancy test results will be listed only. For thyroid function, summaries will be provided by region (Asia vs. Rest of World).

Laboratory results will be summarized/listed using Système International (SI) units, as appropriate. For all quantitative parameters listed in Table 6 , **Error! Reference source not found.**the actual v alue and the change from baseline will be summarized by visit and worst post-baseline visit using descriptive statistics. Summaries will be provided by region (Asia vs. Rest of World). Plots of laboratory values/change from baseline over time will be provided for selected lab parameters. The summary tables will report lab assessments up to 30 days of the last dose date.

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. Summaries will be provided by region (Asia vs. Rest of World).

In addition, Hy's law for liver injuries, hepatic function laboratory test and thyroid laboratory test are summarized.

Table 6: 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, total and direct bilirubin, ALT, AST, alkaline phosphatase, lactate dehydrogenase [LDH] [optional], total protein, albumin, creatine kinase (CK), creatine kinase-cardiac muscle isoenzyme (CK-MB) and amylase, calcium, magnesium and lipase	complete blood count [CBC], including red blood cell [RBC] count, hemoglobin, Hematocrit,white blood cell [WBC] count with automated differential [neutrophils, lymphocytes, and], and platelet count	Prothrombin time (PT),International Normalized Ratio (INR), Partial thromboplastin time or activated partial thromboplastin time	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 HBsAb, HBV DNA and HCV RNA

7.5.4 Vital Signs

Descriptive statistics for vital sign parameters (systolic and diastolic blood pressure [BP], pulse rate, 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.

7.5.5 Ophthalmology Examination

Ophthalmology abnormality will by summarized. In addition ophthalmology examination findings will be listed by patients and visits.

7.5.6 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 by treatment arm.

7.5.7 ECOG

A shift table from baseline to worst post baseline will be provided.

7.6 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 such as population PK analysis may be conducted as appropriate. Exposure – response (efficacy and safety endpoints) analysis may be carried out if supported by data. The results from these analyses will be reported separately from the main study report.

7.7 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-induced 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, and may be reported separately from the main study report.

7.8   OTHER ANALYSIS

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 (collected from patients after confirmed disease progression) if any collected. BOR, ORR, and PFS may be present by biomarker categories. 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.

8   INTERIM ANALYSIS

An interim analysis for OS superiority test will be performed by an independent statistician external to the Sponsor. The independent statistician will work with the blinded study statistician to provide statistical outputs to the IDMC as described in the IDMC charter and perform any ad-hoc analyses requested by the IDMC.

The aim of this interim analysis is to stop the study early for efficacy if there is convincing evidence of outstanding OS benefit. An IDMC will be responsible for making the recommendation regarding stopping the study early based on predefined criteria for OS, as well as results from ORR and other secondary efficacy endpoints. More details will be given in the IDMC charter. Patients will continue to be treated per protocol at the time of interim analysis and until a final decision is made by the Sponsor after considering the recommendation provided by the IDMC.

Timing and Stopping Boundary in the Interim and final Analyses of Overall Survival:

There will be 1 interim analysis of OS utilizing the O'Brien-Fleming boundary approximated by Hwang-Shih-DeCani spending function with the gamma parameter set at -4. The interim analysis

will be performed at the time when approximately 423 death events (87% of the target number of OS events) among the 2 treatment arms are observed. It is estimated that it will take approximately 33 months (from first patient enrolled in study) to observe 423 death events. The IDMC will oversee the interim analysis of OS. The final analysis of OS will take place after approximately 488 OS events have been observed. Stopping boundaries in p-value and Z score for primary analyses of OS are shown in the table below. The boundaries will be updated according to the actual numbers of events in the interim and final analyse using the above pre-specified alpha spending function.

Table 7: **Stopping Boundaries (in p-value and Z score) of Primary Analysis of Overall Survival**

Endpoint	Analysis	Analysis Time (month)	# Events	p-value[1] (Z score) for Efficacy	Approximate HR Threshold
OS	Interim analysis	33	423	< 0.0145 (> 2.18)	0.809
	Final analysis	40	488	< 0.0216 (> 2.02)	0.833

Abbreviations: HR = hazard ratio; OS = overall survival
[1]one-sided

9   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.

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.

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.

Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. New York: John Wiley & Sons; 1980.

11 APPENDIX

11.1 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 > min (death date, data cutoff date, study discontinuation date), then set it as min (death date, data cutoff date, study discontinuation 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. Impute end date first if both start and end dates are partially missing.
If the imputed start date > end date, set to end date.

11.2 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, if subject dies and set to death date, otherwise 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 or end of study date, then set to death date or end of study date.

11.3 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 subsequent anti-cancer therapy is collected from CRF "post-treatment discontinuation anti-cancer systemic therapy" or "post-treatment discontinuation anti-cancer procedure" page, and the **imputed date is prior to the last dosing date, then set to last dosing date + 1**

- If the imputed date > min (death date, data cutoff date, study discontinuation date), then set it as min (death date, data cutoff date, study discontinuation date)

- The imputed start date must be before or equal to the end date

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, data cutoff date), then set to min (death date, study discontinuation date, data cutoff date)

  If year of the start date/stop date is missing, do not impute. Impute partial dates for prior anti-cancer 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.

- 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 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 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.

11.4 RULES FOR IDENTIFYING MISSING TUMOR ASSESSMENTS

Identifying two missing tumor assessment
1) Input scheduled TA visit list for each study
2) every 6 weeks till 48 weeks and every 9 weeks after week 48.Identify last evaluable TA before PD or death (--LPTADT) and map it to the closest scheduled visit (--LPTADT_WK).
   a. In the event of unscheduled TA, choose the closest scheduled visit number (e.g. 6wk or 27wk) as –LPTADT_WK. It can be achieved programmatically by following the classification rule (e.g. defining thresholds) depicted in table 8 below.
   b. Otherwise, assign the scheduled visit number (assuming it is coded correctly) to --LPTADT_WK
3) Find the 2nd TA visit after LPTADT_WK according to the list in step 1 (--LPTADT_WK_2)
   a. If LPTADT_WK_2+1wk < earliest of PD/death date, then censor PFS at the –LPTADT

Table 8 shows how to assign unscheduled TA to a schedule visit. The Threshold column is defined as the mid-point between current and next visit (except for baseline); it is the upper limit for LPTADT to be mapped to the prior scheduled assessment (step 2a above). For example, if LPTADT is Week 44 for an unscheduled visit, it will be mapped to Week 42 TA since it is within the Thredshold for Week 42. Assuming it is SD and the subsequent TA of the patient is PD after Week 58, PFS will be censored at LPTADT (Week 44); had the PD occurred prior to Week 58, it would be counted as an PFS event

Table 8 Example of scheduled tumor assessments with time window

Weeks	Scheduled week -1	**Scheduled week**	Scheduled week+1	Thredshold
Baseline		Baseline		
Every 6 weeks for the first 48 weeks	Week 5	Week6	Week 7	Week 9
	Week 11	Week 12	Week 13	Week 15
	Week 17	Week 18	Week 19	Week 21
	Week 23	Week 24	Week 25	Week 27
	Week 29	Week 30	Week 31	Week 33
	Week 35	Week 36	Week 37	Week 39
	Week 41	Week 42	Week 43	Week 45
	Week 47	**Week 48**	Week 49	Week 52
	Week 56	**Week 57**	Week 58	Week 61

Every 9 weeks afterwards	Week 65	Week 66	Week 67	Week 70
	Week 74	Week 75	Week 76	...

VV-CLIN-092227   Version   2.0

11.5 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

VV-CLIN-092227   Version   2.0