# STATISTICAL ANALYSIS PLAN

| | |
|---|---|
| **Study Protocol Number:** | BGB-A317-305 |
| **Study Protocol Title:** | A Randomized, Double-Blind, Placebo-Controlled, Phase 3 Clinical Study Comparing the Efficacy and Safety of Tislelizumab (BGB-A317) plus Platinum and Fluoropyrimidine Versus Placebo plus Platinum and Fluoropyrimidine as First-Line Treatment in Patients with Locally Advanced Unresectable or Metastatic Gastric or Gastroesophageal Junction Adenocarcinoma |
| **Date:** | Nov 18, 2021 |
| **Version:** | Final version 2.0 |
| **NCT Number:** | NCT03777657 |

## SIGNATURE PAGE

**Author:** [redacted]

[redacted]

## Approval

[redacted]

[redacted]

[redacted]

[redacted]

**TABLE OF CONTENTS**

## DOCUMENT REVISION HISTORY

| Version number | Finalization date | summary of change |
|---|---|---|
| 1.0 | Sept 12, 2021 | |
| 2.0 | Nov 18, 2021 | - Update "TAP score" to more general term "PD-L1+ score" throughout the document. The change does not impact the original PD-L1+ subgroup definition per protocol<br>- Introduction section: Refine wording to introduce TAP<br>- Section 7.1: Replace PD-L1 scores with TAP when defining PD-L1 + analysis set<br>Replace immunogenicity with Antidrug antibody (ADA) when defining ADA analysis set<br>- Section 7.2.5: Clarify that PFS and ORR testing will be conducted based on the date up to interim analysis after positive OS testing in both PD-L1 + analysis set and ITT analysis set (either at interim or final analysis).<br>- Section 7.3.6: Update definition for concomitant medication to align with definition update in TEAE. Delete the summary of systemic Corticosteroid /immunosuppressant<br>- Section 7.4.3: Clarify subgroup analysis for geography and PD-L1 expression<br>- Section 7.4.5 Clarify the new anticancer therapy flags for efficacy and safety analysis<br>- Section 7.5.2: To streamline the logic of deriving TEAE, TEAE definition is updated<br>- Section 7.5.2.1: Clarify that TEAE overview will not include imAE; Overview of TEAE will be summarized by region in safety analysis set Section 7.5.2.2: add sentence to emphasize all imAE up to last dose of tislelizumab/placebo will be summarized, clarity that imAE will be summarize in tislelizumab + chemotherapy arm only |

## 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 |
| BID | Twice daily |
| BIRC | Blinded Independent Review Committee |
| BOR | Best overall response |
| CBC | Complete blood count |
| CBR | Clinical benefit rate |
| CI | Confidence interval |
| CL | Clearance |
| $C_{max}$ | Maximum observed plasma concentration |
| $C_{min}$ | Minimum observed plasma concentration |
| COVID-19 | Coronavirus disease of 2019 |
| CR | Complete response |
| CSR | Clinical Study Report |
| CT | Computed tomography |
| $C_{trough}$ | Trough serum concentration |
| DCR | Disease control rate |
| dMMR | mismatch repair deficient |

| | |
|---|---|
| DOR | Duration of response |
| EBV | Epstein-Barr Virus |
| ECG | Electrocardiogram |
| ECOG PS | Eastern Cooperative Oncology Group Performance Status |
| eCRF | Electronic case report form |
| EDC | Electronic data capture |
| EORTC QLQ-C30 | European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30 |
| EORTC QLQ-STO22 | European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Gastric Cancer Module QLQ-STO22 |
| EQ-5D-5L | European Quality of Life 5-Dimensions 5-Levels Health Questionnaire |
| FDG-PET | Fluorodeoxyglucose-position emission tomography |
| FFPE | Formalin-fixed paraffin-embedded |
| GC | Gastric cancer |
| GCP | Good Clinical Practice |
| GEJ | gastroesophageal junction |
| HR | Hazard ratio |
| HRQoL | Health-related quality of life |
| ICF | Informed consent form |
| ICH | International Conference on Harmonisation |
| iDMC | Independent Data Monitor Committee |
| INR | International Normalized Ratio |
| imAE | Immune-mediated adverse event |

| | |
|---|---|
| ITT | Intent-to-Treat |
| IV | Intravenous |
| IWRS | Interactive Web Response System |
| LLN | Lower limit of normal |
| MedDRA | Medical Dictionary for Regulatory Activities |
| MRI | Magnetic resonance imaging |
| MSI | Microsatellite Instability |
| MTD | Maximum tolerated dose |
| NCI-CTCAE | National Cancer Institute Common Terminology Criteria for Adverse Events |
| NA | Not assessable |
| NE | Not evaluable |
| ORR | Objective response rate |
| OS | Overall survival |
| PD | Progressive disease |
| PD-1 | Programmed cell death protein-1 |
| PD-L1 | programmed cell death ligand-1 |
| PFS | Progression-free survival |
| PK | Pharmacokinetic(s) |
| PO | Orally |
| PP | Per-Protocol |
| PR | Partial response |
| RECIST | Response Evaluation Criteria in Solid Tumors |
| RMST | Restricted Mean Survival Time |

| | |
|---|---|
| RPSFT | Rank Preserving Structural Failure Time |
| ROW | Rest of world |
| Q3W | Once every 3 weeks |
| SAE | Serious adverse event |
| SAP | Statistical Analysis Plan |
| SD | Stable disease |
| SOC | System Organ Class |
| $T_{1/2}$ | Elimination half-life |
| TCM | Tradition Chinese Medicine |
| $T_{max}$ | Time to maximum plasma concentration |
| TA | Tumor assessment |
| TAP | Tumor Area Positivity |
| TEAE | Treatment-emergent adverse event |
| TIC | tumor and immune cell |
| QTcF | Fridericia's correction formula |
| ULN | Upper limit of normal |
| Vd | Volume of distribution |
| 5-FU | 5-fluorouracil |

# 1 INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to describe the procedures and statistical methods that will be used to analyze and report results for study BGB-A317-305: A Randomized, Double-Blind, Placebo-Controlled, Phase 3 Clinical Study Comparing the Efficacy and Safety of Tislelizumab (BGB-A317) plus Platinum and Fluoropyrimidine Versus Placebo plus Platinum and Fluoropyrimidine as First-Line Treatment in Patients with Locally Advanced Unresectable or Metastatic Gastric or Gastroesophageal Junction Adenocarcinoma. This SAP is based on BGB-A317-305 Protocol Amendment 2.0, dated on April 03, 2020. The focus of this SAP is the planned interim and final analysis specified in the study protocol. 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 as needed and will be attached to the clinical study report.

PD-L1 expression is determined by PD-L1 score assessed by tumor area positive score (TAP) (previously referred to Tumor Immune Cell (TIC) score 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.

# 2 STUDY OVERVIEW

## 2.1 STUDY DESIGN

This is a randomized (1:1), double-blind, placebo-controlled, Phase 3 study of tislelizumab plus platinum and fluoropyrimidine versus placebo plus platinum and fluoropyrimidine in patients with locally advanced unresectable or metastatic gastric or GEJ adenocarcinoma.

The study procedures will occur over a Screening Phase (up to 28 days); Treatment Phase (until the investigator determines that none of the study drugs will no longer be used); Safety Follow-up Phase (30 days ± 7 days after the last dose of study drugs, or before the initiation of a new anticancer treatment, whichever occurs first); and a Survival Follow-up Phase (continuing until death, loss to follow-up, withdrawal of consent, or study termination by sponsor).

In addition, immune-mediated AEs will be recorded up to 90 days after the last dose of tislelizumab or placebo, regardless of whether or not the patient starts a new anticancer therapy.

At randomization, patients will be stratified by the following 4 factors:

- Regions of enrollment: China (including Taiwan) vs. Japan and S. Korea vs. US and Europe and other regions. NOTE: other regions include other western countries/populations.
- Programmed cell death protein ligand-1 (PD-L1) expression: positive or negative. PD-L1 positive patients are the ones with PD-L1 score ≥ 5% using VENTANA PD-L1 (SP263) Cdx Assay.
- Presence of peritoneal metastasis (yes or no)

- Investigator's choice of chemotherapy (Oxaliplatin + capecitabine versus cisplatin + 5-fluorouracil [5-FU])

After randomization, patients will be treated on the following arms. Cross-over between the treatment arms will not be allowed.

- Arm A: Tislelizumab 200 mg intravenously (IV) once every 3 weeks (Q3W) + chemotherapy
- Arm B: Placebo IV Q3W + chemotherapy

### 2.2 STUDY ASSESSMENTS

Tumor assessments will be performed by investigator using RECIST v1.1 criteria (Eisenhauer EA 2009). Baseline tumor imaging (computed tomography [CT] with or without contrast or magnetic resonance imaging [MRI]) must be performed within 28 days prior to randomization. On-study tumor assessments will occur every 6 weeks (±7 days) during the first 48 weeks and every 9 weeks (±7 days) thereafter from the randomization until disease progression. If a patient discontinues study treatment due to any reasons other than disease progression, tumor assessments will continue to be performed as scheduled until disease progression, loss to follow up, withdrawal of consent, death, or until the study terminates, whichever occurs first.

Patients will be evaluated for any adverse events (AEs) and serious adverse events (SAEs) occurring up to 30 days after the last dose of study drug (all severity grades), per National Cancer Institute Common Terminology Criteria for Adverse Events [NCI-CTCAE] v5.0 or until initiation of new anticancer therapy, whichever occurs first, and for immune-mediated AEs (imAEs) occurring up to 90 days after the last dose of tislelizumab or placebo regardless of whether or not the patient starts a new anticancer therapy. All drug-related SAEs will be recorded by the investigator after treatment discontinuation until patient death, withdrawal of consent, or loss to follow-up, whichever occurs first. All study drug-related SAEs will be followed until they resolve to baseline or ≤ Grade 1, the investigator assesses the AE as stable and unlikely to improve, or the patient is lost to follow-up, whichever occurs first.

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

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 the IDMC Charter.

PD-L1 status of each individual patient will be blinded to BeiGene study team in order to avoid unwanted bias. PD-L1 expression status in the IRT will not be accessible to anyone except for independent statistician, BeiGene IRT manager, and IDMC. PD-L1 expression status will not be

included in the listings generated for data review either. Study teams from BeiGene and BeiGene's cooperator (e.g.Novartis) can only review the summarized proportions of patients with different PD-L1 expression levels in order to monitor PD-L1 prevalence in the study.

# 3 STUDY OBJECTIVES

## 3.1 PRIMARY OBJECTIVES

- To compare overall survival of tislelizumab plus chemotherapy versus placebo plus chemotherapy in the programmed cell death protein ligand-1 positive (PD-L1+) and intent-to-treat analysis set

## 3.2 SECONDARY OBJECTIVES

- To compare progression-free survival per Response Evaluation Criteria in Solid Tumors 1.1 as assessed by investigators of tislelizumab plus chemotherapy versus placebo plus chemotherapy in the programmed cell death protein ligand-1 positive and intent-to-treat analysis set
- To evaluate overall response rate, and duration of response, per Response Evaluation Criteria in Solid Tumors 1.1 as assessed by investigators
- To evaluate European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Gastric Cancer Module QLQ-STO22 Score, European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30 Score, and European Quality of Life 5-Dimensions 5-Levels Health Questionnaire Score
- To evaluate the safety and tolerability profile of tislelizumab or placebo plus chemotherapy
- To evaluate disease control rate, clinical benefit rate, and time to response per Response Evaluation Criteria in Solid Tumors 1.1 as assessed by investigators

## 3.3 EXPLORATORY OBJECTIVES

- To evaluate progression-free survival after next line of treatment (PFS2)
- To characterize the pharmacokinetics of tislelizumab
- To determine host immunogenicity to tislelizumab
- To assess predictive, prognostic, exploratory biomarkers including but not limited to programmed cell death protein ligand-1 (PD-L1) expression, Epstein-Barr virus (EBV) infection, microsatellite instability-high (MSI-H) or mismatch repair deficient (dMMR) status, genomically stable (GS) or chromosomal instability (CIN), immune-related gene expression profiling, tumor infiltrated lymphocytes (TILs) and tumor mutation burden in tumor tissues and/or blood samples and the association with response to study treatment, mechanisms of resistance, and/or disease status

# 4 DEFINITION OF PRIMARY ESTIMAND

## 4.1 PRIMARY ESTIMAND 1 – SURVIVAL BENEFIT IN GASTRIC OR GEJ ADENOCARCINOMA WITH PD-L1 POSITIVE

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

The primary scientific question of interest is: will the addition of tislelizumab to chemotherapy doublet (platinum and fluropyrimidine) prolong survival in first line gastric or GEJ adenocarcinoma with PD-L1 score ≥ 5%, regardless whether to receive anticancer therapy subsequently.

The primary estimand described by the following attributes:

1. Treatment of interest:

   The experimental treatment regimen constitutes tislelizumab plus chemotherapy (either cisplatin + 5-FU or oxaliplatin + capecitabine). Patients receive capecitabine may be further administrated capecitabine as maintenance therapy. The control treatment regimen is chemotherapy (either cisplatin + 5-FU or oxaliplatin + capecitabine, with capecitabine as maintenance therapy afterwards).

2. Population:

   Adult patients with locally advanced unresectable or metastatic gastric or GEJ adenocarcinoma, who have not received previous systemic therapy for advanced disease and PD-L1 score ≥ 5% using VENTANA PD-L1[SP263] Cdx Assay.

3. Primary variable:

   Overall survival defined as the time from the date of randomization to the date of death due to any cause. Further details on OS are provided in Section 7.4.1.

4. Handling of intercurrent events:

   - New anticancer therapy started prior to death: any incidence will be ignored, i.e., any death or patients' data collected after the new anticancer therapy will be considered for analysis. (treatment policy strategy)
   - Discontinuation of treatment: any death or patients' data collected after the discontinuation of treatment will be considered for analysis. (treatment policy strategy)
   - Death due to COVID-19 infection: it will be counted as an event in the analysis of overall survival. (composite strategy)
   - Treatment interruption/discontinuation due to COVID-19 infection: any incidence will be ignored. Any death or patients' data collected after the treatment interruption/discontinuation due to COVID-19 infection will be considered for analysis. (treatment policy strategy)
   - Any other unforeseen intercurrent events: OS will take into account all deaths and any patients' data after any unforeseen intercurrent events

5. Population-level summary:

Hazard ratio (HR) of OS comparing tislelizumab plus platinum and fluoropyrimidine versus placebo plus platinum and fluoropyrimidine estimated using Cox proportional hazard model stratified by regions of enrollment (east Asia versus rest of the world) and presence of peritoneal metastasis (yes or no).

### 4.2 PRIMARY ESTIMAND 2 – SURVIVAL BENEFIT IN GASTRIC OR GEJ ADENOCARCINOMA

The definition of primary estimand 2 components is similar to the one described in Section 4.1. The application to the specific one in ITT analysis set only is straightforward with minor modification.

The primary scientific question of interest is: will the addition of tislelizumab to chemotherapy doublet (platinum and fluropyrimidine) prolong survival in first line gastric or GEJ adenocarcinoma, regardless whether to receive anticancer therapy subsequently.
The primary estimand described by the following attributes:

1. Treatment of interest:
   The experimental treatment regimen constitutes tislelizumab plus chemotherapy (either cisplatin + 5-FU or oxaliplatin + capecitabine). Patients receive capecitabine may be further administrated capecitabine as maintenance therapy. The control treatment regimen is placebo plus chemotherapy (either cisplatin + 5-FU or oxaliplatin + capecitabine , with capecitabine as maintenance therapy afterwards).

2. Population:
   Adult patients with locally advanced unresectable or metastatic gastric or GEJ adenocarcinoma, who have not received previous systemic therapy for advanced disease.
3. Primary variable:
   Overall survival defined as the time from the date of randomization to the date of death due to any cause. Further details on OS are provided in Section 7.4.1.
4. Handling of remaining intercurrent events:
   - New anticancer therapy started prior to death: any incidence will be ignored, ie. Any death or patients' data collected after the new anticancer therapy will be considered for analysis. (treatment policy)
   - Discontinuation of treatment: any death or patients' data collected after the discontinuation of treatment will be considered for analysis. (treatment policy)
   - Death due to COVID-19 infection will be counted as an event in the analysis of overall survival. (composite strategy)
   - Treatment interruption/discontinuation due to COVID-19 infection: any incidence will be ignored. Any death or patients' data collected after the treatment interruption/discontinuation due to COVID-19 infection will be considered for analysis. (treatment policy)
   - Any other unforeseen intercurrent events: OS will take into account all deaths and any patients' data after any unforeseen intercurrent events

5. Population-level summary:

Hazard ratio (HR) of OS comparing tislelizumab plus platinum and fluoropyrimidine versus placebo plus platinum and fluropyrimidine, estimated using Cox proportional hazard model stratified by regions (east Asia versus rest of the world), presence of peritoneal metastasis (yes or no) and PD-L1 expression (positive: PD-L1 score≥ 5% or negative: PD-L1 score < 5%).

# 5 STUDY ENDPOINTS

## 5.1 PRIMARY ENDPOINTS

- Overall Survival – defined as the time from the date of randomization to the date of death due to any cause

## 5.2 SECONDARY ENDPOINTS

- Progression-free survival as assessed by investigators– defined as the time from the date of randomization to the date of the first objectively documented tumor progression, assessed by investigators per Response Evaluation Criteria in Solid Tumors v1.1, or death, whichever occurs first.
- Overall response rate as assessed by investigators – defined as the proportion of patients whose best overall response is complete response or partial response per Response Evaluation Criteria in Solid Tumors v1.1
- Duration of response as assessed by investigators – defined as the time from the first determination of an objective response per Response Evaluation Criteria in Solid Tumors v1.1, until the first documentation of progression or death, whichever occurs first
- Change from baseline in European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Gastric Cancer Module QLQ-STO22 Score and change from baseline in European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30 Score and European Quality of Life 5-Dimensions 5-Levels Health Questionnaire Score
- The incidence and severity of adverse events according to National Cancer Institute Common Terminology Criteria for Adverse Events v5.0
- Disease control rate (ie, proportion of complete response + partial response + stable disease), clinical benefit rate (ie, proportion of complete response + partial response + durable stable disease) and time to response (ie, time from randomization to the first determination of an objective response) per Response Evaluation Criteria in Solid Tumors 1.1 by investigators

## 5.3 EXPLORATORY ENDPOINTS

- Progression-free survival after next line of treatment (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

- Summary of serum concentration of tislelizumab
- Assessments of immunogenicity of tislelizumab by determining the incidence of antidrug antibodies
- Status of programmed cell death protein ligand-1 (PD-L1) expression, immune or gastric-related, and other exploratory biomarkers including but not limited to Epstein-Barr virus (EBV) infection, microsatellite instability-high (MSI-H) or mismatch repair deficient (dMMR), genomically stable (GS) or chromosomal instability (CIN), immune-related gene expression profiling, tumor infiltrated lymphocytes (TILs) and tumor mutation burden in tumor tissues and/or blood samples obtained before treatment with tislelizumab and/or at progression, and the association with disease status and/or response to tislelizumab in combination with chemotherapy or chemotherapy alone

# 6 SAMPLE SIZE CONSIDERATIONS

The sample size calculation is based on the primary efficacy analyses of OS in the comparison between Arms A and B in the PD-L1 positive (PD-L1+) and ITT analysis sets. The OS is assumed to follow an exponential distribution. Sequential testing procedure is implemented to control overall alpha at 2.5% one-sided. OS analysis in the PD-L1+ analysis set will be performed first and OS analysis in the ITT analysis set will be carried out only if the OS analysis in the PD-L1+ analysis set is statistically significant favoring tislelizumab + chemotherapy arm. Table 1 summarizes the statistical assumption and power in the sample size calculation. Assuming a 50% PD-L1+ (PD-L1score $\geq$ 5%) prevalence rate, a total of 928 patients including approximately 464 (ie, 50%) in the PD-L1+ subset will be enrolled in a 1:1 randomization to observe targeted OS events at the defined time periods as shown in Table1. Assuming a roughly 5% dropout rate, approximately 980 patients will be enrolled over 24 months at enrollment rates of 17 patients/month in the first 2 months, 34 patients/month in the next 2 months, and 44 patients/month in the last 20 months. The enrollment assumptions, including percentage of PD-L1+ patients, will be monitored during the enrollment; therefore, the sample size and timeline could be adjusted accordingly. Enrollment of patients whose tumors are PD-L1- might be stopped if necessary, to ensure that the percentage of PD-L1+ (PD-L1 score $\geq$ 5%) is no less than 50% of the ITT analysis set. The primary analyses will be performed when the target number of events in both ITT and PD-L1+ analysis set are observed. An interim analysis of OS is planned after approximately 70% of the total planned death events have occurred in both ITT and PD-L1+ analysis sets (Section 8 for more details).

**Table 1: Hazard ratio and median OS assumption, number of events, alpha and power in the primary hypothesis tests**

| Analysis Set | HR | Median in Arm A (in months) | Median in Arm B (in months) | # Events | Alpha | Power |
|---|---|---|---|---|---|---|
| PD-L1+ | 0.75 | 15.3 | 11.5 | 384 | 0.025 | 80% |
| ITT | 0.8 | 14.4 | 11.5 | 768 | 0.025 | 87% |

# 7 STATISTICAL METHODS

## 7.1 ANALYSIS POPULATIONS

Intent-to-Treat (ITT) analysis set – Includes all randomized patients. Patients will be analyzed according to their treatment assigned at randomization. This will be one of the dual primary analysis sets for demography and efficacy analyses.

PD-L1 positive analysis set (PD-L1 score ≥ 5% using VENTANA PD-L1 (SP263) assay). – Includes all randomized patients whose tumors were PD-L1 positive (PD-L1 score ≥ 5%). Patients will be analyzed according to their treatment assigned at randomization. This will be one of the dual primary analysis sets for demography and efficacy analyses.

Safety analysis set – Includes all patients who received at least 1 dose of study drugs. This will be the analysis set for the safety analyses. Patients will be analyzed according to the study treatment they actually received. Patients will be classified according to treatment received, where treatment received is defined as (i) the intended treatment if it was received at least once, or (ii) the first treatment received when starting therapy with study medication if intended treatment is never received. Each patient will be classified into and analyzed consistently within one (and only one) treatment arm.

PD-L1 positive safety analysis set – Includes patients in the safety analysis set whose PD-L1 score ≥ 5% using VENTANA PD L1 [SP263] Cdx Assay. This will be analysis set for the safety analyses in patients with PD-L1 score≥ 5%.

PK Analysis Set included all patients who received ≥ 1 dose of tislelizumab per the protocol for whom any post dose PK data were available.

Antidrug antibody (ADA) Analysis Set included all patients who received ≥ 1 dose of tislelizumab for whom both baseline ADA and ≥ 1 postbaseline ADA results were available.

## 7.2 DATA ANALYSIS GENERAL CONSIDERATIONS

### 7.2.1 Definitions and Computations

Study day:

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

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

Baseline Measurements:

- For efficacy evaluation: a baseline value is defined as the last non-missing value collected prior to the randomization.

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

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

Minimum study follow up is defined as a difference between the date of analysis cut-off and the date of last patient randomized.

All calculations and analyses will be conducted using SAS version 9.2 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 when 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/partial Data

Handling of missing data related to primary estimand will be further elaborated in Section 7.4.1. Missing dates or partially missing dates will be imputed conservatively for adverse events and prior/concomitant medications/procedures. Specific rules for handling of missing or partially

missing dates for adverse events and prior/concomitant medications/procedures are provided in Section 11 Appendix 1. Other missing data will not be imputed unless otherwise specified in the SAP.

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

### 7.2.4 Adjustments for Covariates/Stratification

The value of the stratification factors used at randomization (IVRS), including region of enrollment, PD-L1 expression [in ITT only] and presence of peritoneal metastasis will be used in stratified log-rank test and stratified Cox proportional hazard model for primary endpoint OS, secondary endpoint, PFS. Similarly these stratification factors will be used in Cochran-Mantel-Haenszel method to analyze ORR.

The actual value of the stratification factors (collected in eCRF) and other baseline covariates may be used in statistical models as covariates as sensitivity analyses for endpoints.

### 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 PD-L1+ and ITT analysis sets using sequential testing method. OS analysis in ITT analysis set will be performed only if the OS analysis in the PD-L1 positive analysis set is statistically significant favoring tislelizumab + chemotherapy arm.

Hypothesis testing of the secondary endpoints PFS and ORR in the PD-L1 positive and ITT analysis sets will be performed at the same time as the interim analysis of OS. Only when the superiority of OS in both PD-L1+ and ITT analysis sets have been demonstrated (either at interim or final analysis), full alpha of 0.025 (1-sided) will be shifted sequentially to the hypothesis testing of the secondary endpoints in order of PFS in PD-L1+, ORR in the PD-L1+ analysis set, followed by PFS in ITT and ORR in ITT analysis set based on the data up to interim analysis. The hypothesis test will be stopped at the first non-significant endpoint. Nominal p-values may be computed for other efficacy analysis but should be interpreted with caution.

Inferential tests of efficacy outcomes will be carried out as described above. Type I error is not controlled in the safety analysis and safety summary will be performed in the safety analysis set regardless of the results in primary efficacy endpoint testing, unless otherwise specified.

### 7.2.6 Data Integrity

The data set for analysis should be an accurate and complete representation of the patients' relevant outcomes from the clinical database. All data should be complete and reviewed up to a pre-specified cutoff date as specified in the Data Extract and Snapshot Plan. Consistency checks and appropriate source data verification should be completed as specified in the Site Monitoring Plan.

## 7.3 SUBJECT CHARACTERISTICS

### 7.3.1 Subject Disposition

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

The number (percentage) of patients randomized, treated, discontinued from treatment (or completed chemotherapies) and discontinued from the study will be summarized. The primary reason for end of treatment (treatment discontinuation) and end of study (study discontinuation) will be summarized by categories. The reasons for treatment/study discontinuation related to COVID-19 impact will also be summarized. Study follow up duration will be summarized descriptively.

Patient disposition will also be summarized by region (China (including Taiwan) vs. Japan and S. Korea vs. US and Europe and other regions) for ITT and PD-L1 positive analysis set.

### 7.3.2 Protocol Deviations

Protocol deviation criteria will be established together with its category/term of important and 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.

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

Protocol deviations that are related to COVID-19 will be summarized and all patients affected by COVID-19-related protocol deviations will be listed along with site information and specific deviation description.

### 7.3.3 Demographic and Other Baseline Characteristics

Demographic and other baseline characteristics will be summarized in the ITT and PD-L1 positive analysis set using descriptive statistics. Continuous variables will be summarized using number of patients, mean, standard deviation, median, minimum and maximum. Categorical variables will be summarized using number of patients and percentage in relevant categories.

Demographic and other baseline characteristics include:

- Age
- Age group (≤ 65 vs > 65 years)
- Sex
- Race
- Ethnicity
- Weight (kg)
- BMI
- ECOG

- Smoking status
- Alcohol consumption

In addition, the stratification factors per Interactive Response Technology (IRT) and per eCRF will be summarized based on ITT population:

- Regions of enrollment: China (including Taiwan) vs. Japan and S. Korea vs. US and Europe and other regions.
- PD-L1 expression (PD-L1 score ≥ 5% vs PD-L1 score < 5%):
- Presence of peritoneal metastasis (yes or no)
- Investigator's choice of chemotherapy (Oxaliplatin + capecitabine versus cisplatin + 5 fluorouracil [5-FU])

### 7.3.4 Disease History and Baseline Disease Characteristics

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

- Time since initial cancer diagnosis to study entry
- Disease stage at screening
- Location of tumor
- PD-L1 expression (<1%, ≥1%, ≥5% and ≥10%)
- HER2 status
- Histological Tumor differentiation grade
- Histologic type (Lauren classification)
- MSI or MMR status
- EBV status
- Genomically stable type
- Chromosomal instability type
- Patients with non-target lesion only
- Prior gastrostomy/ esophagectomy
- Previous adjuvant or neoadjuvant therapy
- Number of metastatic sites, 0-2 vs >=3

Disease history and baseline disease characteristics will also be summarized by region (China, Japan & South Korea and ROW) for ITT and PD-L1 positive analysis set. In addition, Disease history and baseline disease characteristics will also be summarized in all randomized patients with PD-L1 score < 5%.

### 7.3.5 Prior Anti-Cancer Therapies and Surgeries

The number of patients receiving prior anti-cancer drug therapies, prior anti-cancer radiotherapy, prior anti-cancer surgery will be summarized. The therapies with the same sequence/regimen number are counted as one prior therapy.

### 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 patients 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 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.

Medication to treat COVID-19 related adverse events will also be summarized separately.

Patient data listings of prior and concomitant medication will be provided.

### 7.3.7 Medical History

Medical History will be coded Medical history/current medical conditions are coded using the latest version of Medical Dictionary for Regulatory Activities (MedDRA) terminology available at the time of the analyses. The number (percentage) of patients reporting a history of any medical condition, as recorded on the CRF, will be summarized by system organ class and preferred term in the safety population. A listing of medical history will be provided.

## 7.4 EFFICACY ANALYSIS

OS is the primary endpoint in this study. OS analysis will be carried out in the PDL1+ analysis set first. A formal statistical test of OS in the ITT analysis set will be performed only if the OS analysis in the PD-L1+ analysis set is statistically significant favoring tislelizumab + chemotherapy. There is one interim analysis of OS for both efficacy and futility planned. The interim analysis of OS will be performed when approximately 269 deaths in the PD-L1+ analysis set and 538 deaths in the ITT analysis set (70% of the target number of OS events in each analysis set) among the 2 treatment arms have been observed. The final analysis of OS will take place after approximately 384 and 768 death events have been observed in the 2 analysis sets,

### 7.4.1 Primary Efficacy Endpoints

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

*Variable*

Overall survival is defined as 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 cut-off date whichever comes first for other alive patients.

Every effort should be made to ensure complete death dates. In the rare case, if day of death date is missing, death date is imputed as the max (last available date showing patients alive + 1, first day of year/month of death date). The patient with imputed death date will be considered as an event for OS analysis.

*Primary efficacy analyses*

OS in PD-L1+ and ITT analysis sets:

The null hypothesis to be tested is:

$H_0$: HR $\geq$ 1

against the alternative:

$H_1$: HR < 1

The primary efficacy analyses of OS will be performed using the stratified log-rank test on the PD-L1+ analysis set and ITT analysis set separately according to the treatment arm and randomization strata (region of enrollment, PD-L1 expression [in ITT only] and presence of peritoneal metastasis) to which patients are assigned at randomization.

The survival distribution of OS will be estimated using Kaplan-Meier method. The results will be plotted graphically (Kaplan-Meier curves) by treatment. The plots will display the number of patients at risk at equidistant time point. The median OS for each treatment group will be provided along with the approximate 95% confidence intervals using the method of Brookmeyer and Crowley (Brookmeyer & Crowley, 1982). Additionally, 25% and 75% percentiles will also be provided. The cumulative probability of OS at every 6 months if estimable, will be calculated for each treatment arm and presented with 2--sided 95% CIs. The CIs will be constructed using the standard errors derived from Greenwood's formula (Collett 1994).

The treatment effect will be estimated by fitting a stratified Cox proportional hazard model to the OS times including treatment arm as a covariate and region of enrollment (east Asia versus ROW), PD-L1 expression (for ITT only) and presence of peritoneal metastasis as strata. From this model, the hazard ratio (HR) of OS will be estimated and presented with a 2-sided 95% CI. In this analysis the baseline hazard function will be allowed to vary across strata. SAS PHREG procedure with TIES=Efron option will be used to carry out this analysis in which the model statement will include treatment group variable as the only covariate and the STRATA statement will include the stratification variable as obtained via IRT.

***Sensitivity analysis and supportive analysis***

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

Sensitivity analysis 2: "OS analysis with stratification factors from the eCRF": OS will be analyzed using stratified Cox model and the distribution of OS will be compared between the treatment group using stratified log-rank test, with stratification factors obtained from the CRF.

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

***Supplementary analysis***

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

Supplementary analysis 2 "OS analysis based on Restricted mean survival time method": This analysis targets an estimand which has the same attributes as the primary estimand except the population level summary 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 separately using the area under the curve from baseline to the minimum of the largest observed time on each of the two treatment groups. RMST will be computed for each treatment arm and the difference with its 95% CI will be displayed.

Supplementary analysis 3 "OS analysis adjusted for baseline covariates": OS will be also analyzed by adjusting multivariate covariates at baseline. The analysis addresses a different scientific question. i.e., will the addition of tislelizumab to chemotherapy doublet prolong survival, adjusting for covariates (ECOG PS, liver metastasis, number of metastatic organs, prior gastrostomy) Covariate adjusted multivariate Cox regression provides a conditional treatment effect rather than marginal treatment effect. A stratified Cox regression model will be performed adjusted key baseline prognostic factors (ECOG PS, live metastasis, number of metastatic organs, prior gastrostomy).

Supplementary analysis 4 "COVID-19 OS supplementary analysis": this analysis will aim at assessing the treatment effect based on OS had COVID-19 pandemic not occurred. 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 strategy)
- **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.

The discontinuation reason due to COVID-19 pandemic will be identified from the eCRF (hypothetical strategy).

- **Medications used for treating COVID-19 cases:** OS will be censored on the date of administration of COVID-19 medication (hypothetical strategy)
- **Death due to COVID-19**: OS will be censored on the date of death due to COVID-19 (hypothetical strategy)

Additional supplementary analyses may be carried out to explore the impact of next-line anticancer therapy to OS evaluation.

### 7.4.2 Secondary Efficacy Endpoints

Progression Free Survival (PFS) by investigators

PFS is defined as the time from the randomization date to disease progression or death, whichever occurs first. PFS will be censored at the last adequate tumor assessment if one of the following occurs: absence of event; the event occurred after a new anticancer therapy is given; the event occurred after two or more consecutive missing/non-evaluable tumor assessments. The table 2 shows the derivation rules for PFS. The algorithm to identify missing TAs are presented in Appendix 2.

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

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

The methods used to analyze OS will be applied to the analysis for PFS. The primary analysis of comparing PFS as assessed by investigators between two treatment groups in the PD-L1 analysis set will be stratified log-rank test, stratified by pooled stratification factors of regions of enrollment

(east Asia versus rest of the world [ROW]), presence of peritoneal metastasis. The primary analysis of comparing PFS as assessed by investigators between two treatment groups in the ITT analysis set will be stratified log-rank test, stratified by pooled stratification factors of regions of enrollment (east Asia versus rest of the world [ROW]), presence of peritoneal metastasis and PD-L1 expression.

The median PFS and the cumulative probability of PFS estimated at every 3 months will be calculated using Kaplan-Meier estimates for each treatment arm and presented with 2-sided 95% CIs computed by Brookmeyer and Crowley method using the log-log transformation. The treatment effect will be estimated by fitting a Cox regression model to the PFS times including treatment arm as a covariate and beforementioned stratification factors as strata. From this model, the hazard ratio (HR) of PFS will be estimated and presented with a 2-sided 95% CI.

The following sensitivity analyses are planned:

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

One supplementary analysis is planned:

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

Objective Response Rate by investigators

Best overall response (BOR), 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. ORR is defined as the number of patients whose BOR is confirmed CR or PR divided by the number of randomized patients in each arm. The null hypotheses of no difference in ORR per RECIST 1.1 assessed by investigators will be tested in a Cochran-Mantel-Haenszel (CMH) test adjusting for region of enrollment (east Asia versus ROW), PD-L1 expression (for ITT only) and presence of peritoneal metastasis as strata in the PDL1+ and ITT analysis sets. Patients with no post-baseline response assessment (for any reason) will be considered as non-responders. The 2-sided 95% CIs for the odds ratio in ORR will be calculated, as well as Clopper-Pearson 95% CIs for ORR and its corresponding Clopper-Pearson 95% CI for each of the response categories (CR, PR, SD, and PD) will be presented by treatment arm.

Duration of Response

Duration of Response (DOR) is defined as progression/death event free time counted from the first objective response date to the first documented radiological PD date/or death date, whichever occurred first. All the censoring rules for PFS should be applied to DOR. Duration of response (DOR) assessed by investigators will be analyzed in the responders only. The median DOR and

the cumulative probability of DOR estimated at every 3 months will be calculated using Kaplan-Meier estimates for each treatment arm and presented with 2-sided 95% CIs computed by Brookmeyer and Crowley method using the log-log transformation. No formal testing will be performed to compare DOR between two treatment group as it would be based on a non-randomized subgroup.

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

Disease control rate (DCR) defined as the proportion of patients whose best overall response (BOR) is CR, PR, or SD including non-CR/non-PD. Clinical benefit rate (CBR) defined as the proportion of patients who have CR, PR, or SD including non-CR/non-PD of $\geq 24$ weeks in duration. DCR and clinical benefit rate (CBR) assessed by investigators will be analyzed similarly to ORR in the ITT and PD-L1+ analysis sets.

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

Waterfall plots will be provided for the maximum tumor shrinkage based on target lesion. In addition, patients will be marked out in the plot for those who had tumor reduction in target lesion assessments but contradicted to the PD results in overall response due to new-lesion or non-target lesion. The maximum tumor shrinkage based on target lesion used in the plots will be listed. These analyses will be performed based on RECIST1.1.

Health-Related Quality of Life

The EORTC QLQ-C30 (QlQ-C-30) consists of thirty questions that are specific to cancer and cancer treatment (Aaronson NK, et al., 1993; Fayers PM, et al., 2001). It includes global health status (GHS/Qol) scale that consists of 2 items, and five functional scales measuring Physical (5 items), Role (2 items), Cognitive (2 items), Emotional (4 items), and Social (2 items), three symptom scales measuring Fatigue (3 items), Pain (2 items), and Nausea and Vomiting (2 items) and six single items measuring Dyspnoea, Insomnia, Appetite Loss, Constipation, Diarrhea, and Financial Difficulties. QLQ-C30 scores are based on 4-point Likert scales from 1 = "Not At All" to 4 = "Very Much" with lower score indicating better HRQoL; except for the GHS/Qol that is scored on a 7-point scale ranging from 1 = "Very Poor" to 7 = "Excellent" with higher scores indicating better health status. The QLQ-STO22 consists of 22 questions in 6 subscales: Dysphagia/odynophagia (4 items), Pain/discomfort (3 items), Dietary restrictions (5 items), Upper gastro-intestinal symptoms (3 items), Specific emotional problems (3 items) and 4 single items. Scores are based on 4-level Likert scales (0= Not at all to 4=Very Much); with lower scores indicating lower symptoms/better HRQoL.

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

**EORTC Scoring Derivation**

The principle for scoring applies to all scales/scores. The derived scales are obtained from the raw scores as defined in the EORTC manual. If at least half of the items for a scale are answered, then the remaining completed items are used to calculate the score for that scale; however, if more than half are missing, the scale score is set to missing.

**Raw Score (RS)**

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

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

**Derived Scale (S)**

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

Functional scales:

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

Symptom scales and global health status:

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

QLQ-STO22 Index score = [∑(domain scores + single item scores)]÷number of available domains and single items

**Scales/Items: QLQ-C30**

GHS and Functional Scales: Higher scores = Better HRQoL

Symptom Scales: Lower Scores = Better HRQoL

**Table 3 QLQ-30 scores**

| | Scale | Number of items | Item range | Item Numbers |
|---|---|---|---|---|
| **Global health status/ QoL**<br>Global health status/QOL | QL2 | 2 | 6 | 29,30 |
| **Functional Scales** | | | | |
| Physical functioning | PF2 | 5 | 3 | 1, 2, 3, 4, 5 |
| Role functioning | RF2 | 2 | 3 | 6, 7 |
| Emotional functioning | EF | 4 | 3 | 21, 22, 23, 24 |

| | | | | |
|---|---|---|---|---|
| Cognitive functioning | CF | 2 | 3 | 20, 25 |
| Social functioning | SF | 2 | 3 | 26, 27 |
| **Symptom Scales/ items** | | | | |
| Fatigue | FA | 3 | 3 | 10, 12, 18 |
| Nausea and vomiting | NV | 2 | 3 | 14, 15 |
| Pain* | PA | 2 | 3 | 9, 19 |
| Dyspnoea | DY | 1 | 3 | 8 |
| Insomnia | SL | 1 | 3 | 11 |
| Appetite loss | AP | 1 | 3 | 13 |
| Constipation | CO | 1 | 3 | 16 |
| Diarrhoea | DI | 1 | 3 | 17 |
| Financial Difficulties | FI | 1 | 3 | 28 |

*questions regarding Pain in QlQ-C30 are more general whereas for HCC & LC13 the pain questions are very specific to the disease. In some of the analysis, there were not significant correlations between them, and in some other studies they were. For simplicity, I suggest for the 317-208 and 307 using the disease specific pain scales; however, for the 311-302 using pain scale from the C-30 scale.

**QLQ – STO22**

Lower scores = Better HRQoL

## Table 4 QLQ-STO22 scores

| Symptom Scales | Number of items | Item range | Item Numbers |
|---|---|---|---|
| Scales | | | |
| Dysphagia/odynophagia | 4 | 4 | 31, 32, 33, 34 |
| Pain/discomfort | 3 | 4 | 35, 36, 37 |
| Upper gastro-intestinal symptoms | 3 | 4 | 38, 39, 40 |
| Dietary restrictions | 5 | 4 | 41, 42, 43, 45, 46 |
| Anxiety/ emotional problems | 3 | 4 | 47, 49, 50 |
| Hair Loss | 2 | 4 | 51, 52 |

| | | | |
|---|---|---|---|
| Single items | | | |
| Dry Month | 1 | | 44 |
| Taste problem | 1 | | 45 |
| Body image | 1 | | 49 |

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

Any unscheduled QoL assessments will be mapped into the scheduled QoL visits (screening or baseline, every cycle up to cycle 6, then every other cycle thereafter) prior to or after this visit if no QoL assessments are available at the scheduled QoL visits.

A questionnaire module is considered complete if at least one question is answered. The adjusted completion rate is defined as percentages of patients who completed the questionnaire at each visit divided by the number of patients still in treatment who were expected to complete the questionnaire. Completion rates and adjusted completion rate for the EORTC QLQ C30, EORTC QLQ -STO22, and EQ-5D-5L will be summarized separately at each visit.

The derived scores (functional scales/symptom scales/single items and the GHS/QoL scale) of QLQ C30, and index and symptom scales/items of QLQ-STO22 will be summarized as well as change from baseline at each cycle using descriptive statistics. For EQ-5D-5L, descriptive modules will be summarized by visit and dimension in an ordinal scale. The plots of the mean values and their standard errors at each visit over time for each sub domain and overall QLQ-STO22 and EQ-5D-5L by treatment arm will be presented.

Time to deterioration analysis will be performed to the compare key quality of life scores between two treatment groups. Time to deterioration is defined as time from randomization to the first occurrence of an decrease of ≥10 scores in global health status/QoL (GHS/QoL) and physical function and fatigue of the QLQ-C30, and ≥10 increase in the index score and scores of dysphagia, pain, dietary restrictions and upper gastro of QLQ-STO22. The primary definition of 10 points is derived from the published 10-point EORTC threshold (Osaoba 1992, Osaba 2002, Osaba et al 1998). Sensitivity analysis may explore a secondary deterioration threshold defined using anchor based meaningful-within patient change (MWPC) framework. A deterioration is not counted as an event if a subsequent improvement returned the overall worsening from baseline to less than 10 points. If a patient does not have an event (10% deterioration), they are censored at their last clinic visit at which HRQoL is measured. A nonparametric Kaplan-Meier method will be used to estimate the deterioration curve in each group. The log-rank test and hazard ratio will be provided to show the magnitude of treatment effect and are only used for descriptive purpose only.

In addition, a mixed effect model analysis for measuring clinically meaningful changes post-baseline will be performed using the GHS/QoL, physical function and fatigue domains of QLQ-C30, and dysphagia, pain, dietary restrictions and upper gastro of QLQ-STO22. Difference in change from baseline to cycle 4 and cycle 6 between arms of before-mentioned parameters will be assessed in the mixed models which include baseline score, stratification factors, treatment arm, visit, and treatment arm by visit interaction as fixed effects and visit as a repeated measure with an unstructured covariance structure.

### 7.4.3 Subgroup Analyses

To determine if the treatment effect is consistent across various subgroups, the median OS in each subgroups along with unstratified OS hazard ratio and its 95% CI will be estimated and plotted within each category of the following variables:

- geography 1 (Asia vs Rest of World)
- geography 2 (China (including Taiwan vs Japan and S. Korea vs Rest of World)
- race
- age (< 65 vs ≥ 65 years)
- gender (Female vs Male)
- ECOG PS (0 vs 1)
- Liver metastasis
- MSI or MMR status
- Presence of peritoneal metastasis (yes or no)
- Investigator's choice of chemotherapy (Oxaliplatin + capecitabine versus cisplatin + 5 fluorouracil [5-FU])
- Prior gastrostomy
- Previous adjuvant or neoadjuvant therapy
- Number of metastatic sites, 0-2 vs >=3
- PD-L1 expression (<5%, ≥5%)

Country-specific subgroups may also be summarized per local regulatory requirements. Subgroup analysis of PD-L1 expression (<1%, ≥1% and < 10%, ≥10%) maybe explored.

### 7.4.4 Exploratory Efficacy Endpoints

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

- 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

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

### 7.4.5 Post and during-treatment Anti-Cancer Therapy

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

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

- As for efficacy analysis, start date of new anti-cancer therapy will be the earlist date of prohibited anti-cancer therapy taken during treatment, date of the post-treatment systemic anti-cancer therapy and date of other anti-cancer therapy such as post-treatment surgery and radiotherapy as deemed appropriate.
- The start date of new anti-cancer therapy in defining TEAE for safety 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 TCM. ORR, PFS or OS benefit of Chinese herbal medicines or Chinese patent medicines has not yet been established. Therefore, they will not be taken into account as new anti-cancer therapy in the efficacy and safety analyses.

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

## 7.5 SAFETY ANALYSES

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

### 7.5.1 Extent of Exposure

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

Treatment duration (TD) for tislelizumab/placebo, cisplatin, 5-FU, oxaliplatin, capecitabine: The treatment duration will be calculated as (last date of exposure – date of first dose + 1)

- If patients discontinued treatment (with non-missing EOT date), using min (CUOFFDT, death date, last dose date + 20) as the "last date of exposure" for tislelizumab, cisplatin and oxaliplatin; using min (CUOFFDT, death date, last dose date + 16) as the "last date of exposure" for 5-FU; using min (CUOFFDT, death date, last dose date + 7) as the "last date of exposure" for capecitabine.
- otherwise if patient has treatment ongoing, using cutoff date as the "last date of exposure" for calculation of TD

Total Cumulative Dose: sum (all actual dosages per administration at all visits prior to the cutoff date)

Actual dose intensity (ADI), planned dose and relative dose intensity (RDI)
Actual Dose Intensity (ADI) for tislelizumab (mg/cycle) = 21*total cumulative dose (mg) / (last dose date prior to cut off date+ 21 - first dose date).

Planned Dose Intensity for tislelizumab (mg/cycle) = 200 mg/cycle.

Relative Dose Intensity (%) for tislelizumab = Actual Dose Intensity/200 *100%

The derivations of ADI planned dose and RDI for Chemotherapy are shown in Table 5.

**Table 5 ADI, planned dose and RDI for Chemotherapy**

| | ADI(mg/m$^2$/cycle) | Planned dose per cycle | RDI |
|---|---|---|---|
| cisplatin | $\frac{\sum_1^{\#of\ cycles} \frac{actual\ dose}{BSA*} \times 21}{date\ of\ last\ dose\ up\ to\ cutoff + 21 - first\ dose\ date}$ | 80 mg/m$^2$ | $\frac{ADI}{80}$ |
| Oxaliplatin | $\frac{\sum_1^{\#of\ cycles} \frac{actual\ dose}{BSA*} \times 21}{date\ of\ last\ dose\ up\ to\ cutoff + 21 - first\ dose\ date}$ | 130 mg/m$^2$ | $\frac{ADI}{130}$ |
| 5-FU | $\frac{\sum_1^{\#of\ cycles} \frac{actual\ dose}{BSA*}}{\max(\frac{date\ of\ last\ dose\ up\ to\ cutoff + 17 - first\ dose\ date}{21}, number\ of\ cycles\ in\ last\ dosing\ CRF\ page)}$ | 800 * 5 mg/m$^2$ | $\frac{ADI}{4000}$ |
| capecitabine | $\frac{\sum_1^{\#of\ cycles} \frac{actual\ dose}{BSA*}}{\max(\frac{date\ of\ last\ dose\ up\ to\ cutoff + 8 - first\ dose\ date}{21}, number\ of\ cycles\ in\ last\ dosing\ CRF\ page)}$ | 2000*14 mg/m$^2$ | $\frac{ADI}{28000}$ |

The number of patients with dose reductions, dose delays, dose interruptions, and treatment discontinuation and their reasons will be summarized by counts and percentages according to study drug.

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

### 7.5.2 Adverse Events

The AE verbatim descriptions (Investigator's description from the eCRF) will be classified into standardized medical terminology using Medical Dictionary for Regulatory Activities (MedDRA). Adverse events will be coded to MedDRA (Version 24 or higher) lower level term closest to the verbatim term. The linked MedDRA System Organ Class (SOC) and Preferred Term are also classified. All adverse event summaries are based on safety analysis set.

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

The incidence of TEAEs will be reported as the number (percentage) of patients with TEAEs by SOC and Preferred Term. A patient will be counted only once by the highest severity grade per NCI-CTCAEv.5.0 within an SOC and Preferred Term, even if the patient experienced more than 1 TEAE within a specific SOC and Preferred Term. The number (percentage) of patients with TEAEs will also be summarized by relationship to the study drug.

#### 7.5.2.1 Treatment Emergent Adverse Event

An overall summary of TEAEs will summarize (in the safety analysis set and PD-L1+ safety analysis set, as well as by region) the number (%) of patients with

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

- At least one TEAE leading to dose modification of study drug
- At least one infusion-related reaction
- At least one infusion-related reaction with NCI-CTCAE grade ≥3

Summaries of the following TEAEs will be provided in the PD-L1+ analysis set and safety analysis set:

- All TEAEs
    - All TEAEs by SOC
    - All TEAEs by SOC and PT
    - Most frequently reported (incidence ≥10% in any treatment arm) TEAE by SOC and PT
    - Treatment-related TEAE by SOC and PT
    - Most frequently reported (incidence ≥10% in any treatment arm) Treatment-related TEAE by SOC and PT
- Serious TEAEs by SOC and PT
    - Most frequently reported (incidence ≥ 5% in any treatment arm) serious TEAE by SOC and PT
    - Treatment-related Serious TEAE by SOC and PT
- TEAEs with NCI-CTCAE grade ≥3 by SOC and PT
    - Treatment-related TEAE with NCI-CTCAE grade ≥ 3 by SOC and PT
- Most frequently reported (incidence ≥ 5 % in any treatment arm) TEAE with NCI-CTCAE grade ≥ 3 by SOC and PT.
- TEAEs leading to death by SOC and PT
    - Treatment-related TEAE Leading to Death by SOC and PT
- TEAEs leading to treatment discontinuation by SOC and PT
    - Treatment-related TEAE Leading to Treatment Discontinuation by SOC and PT
- TEAEs leading to dose modification by SOC and PT
    - Treatment-related TEAE Leading to Dose Modification by SOC and PT

#### 7.5.2.2 Immune-mediated Adverse Event

Immune-mediated adverse events are of special interest and summarized by category within a pre-defined list. The identification of immune-mediated adverse events is described in immune-mediated adverse event charter. All imAE up to 90 days of last dose of tislelizumab/placebo will be summarized.

Summaries of the following incidence of immune-mediated adverse events will be provided in tislelizumab + chemotherapy arm:

- Immune-mediated adverse events by category and maximum severity
- Immune-mediated adverse events with NCI-CTCAE grade ≥3 by category
- 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
- Immune-mediated adverse events treated with systematic corticosteroid by category
- Time to Immune-mediated adverse events and resolution
- Duration of resolved Immune-mediated adverse events

#### 7.5.2.3 Infusion-related Adverse Event

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

### 7.5.3 Death

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

### 7.5.4 Laboratory Values

Clinical laboratory (e.g., hematology, serum chemistry, Coagulation and Urinalysis) values will be evaluated for each laboratory parameter by patient. Abnormal laboratory values will be flagged and identified as those outside (above or below) the normal range. Reference (normal) ranges for laboratory parameters will be included in the CSR for this protocol. Descriptive summary statistics (e.g., n, mean, standard deviation, median, minimum, maximum for continuous variables; n [%] for categorical variables) for laboratory parameters and their changes from baseline will be calculated. Laboratory values will be summarized by visit and by worst postbaseline visit. The summary tables will report lab assessments up to 30 days of the last dose date.

Laboratory parameters that are graded in NCI-CTCAEv.5.0 will be summarized by NCI-CTCAE grade. In the summary of laboratory parameters by NCI-CTCAE grade, parameters with NCI-CTCAE grading in both high and low directions (e.g., calcium, glucose, magnesium, potassium, sodium) will be summarized separately.

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

### 7.5.5 Vital Signs

Descriptive statistics for vital sign parameters (systolic and diastolic BP, heart rate, temperature, 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 height and weight. Vital signs will be listed by patient and visit.

### 7.5.6 Electrocardiograms (ECG)

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

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

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

### 7.5.7 ECOG

ECOG performance status will be summarized by treatment arm and by visit.

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

Patient listing of ECOG will be provided for all ECOG findings.

### 7.5.8 Antidrug Antibody

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

### 7.5.9 Other Safety Measurements

Other safety measurements including pulmonary function, ophthalmology exam, digital pulse oximetry, and abnormal physical exam findings, will be listed by patient and visit.

## 7.6 PHARMACOKINETIC ANALYSES

Pharmacokinetic samples will be collected in this study as outlined in appendix 1 of the protocol, and only from patients randomized to receive tislelizumab in sites that are able to adequately perform PK sampling, handling, and processing procedures as outlined in the Laboratory Manual.

Tislelizumab post-dose and trough serum concentration data (Ctrough) will be tabulated and summarized by visit/cycle at which these concentrations are collected. Descriptive statistics will include means, medians, ranges, and standard deviations, as appropriate. PK analysis by subgroup (e.g. region) may be performed as needed.

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

## 7.7 OTHER ANALYSIS

Distribution of PD-L1 expression will be examined in the ITT Population. Potential association between PD-L1 expression and tislelizumab treatment combined with chemotherapy effect over chemotherapy will be explored.

Other potential predictive markers may be assessed.

# 8 INTERIM ANALYSIS

There will be one interim analysis of OS using the O'Brien-Fleming boundary approximated by Hwang-Shih-DeCani spending function with the gamma parameter set at -4. The non-binding lower (futility) boundary is defined by Hwang-Shih-DeCani spending function with the gamma parameter set at -12. The interim analysis of OS will be performed when approximately 269 deaths in the PD-L1+ analysis set and 538 deaths in the ITT analysis set (70% of the target number of OS events in each analysis set) among the 2 treatment arms have been observed which is estimated to occur approximately 30 months after the first patient is randomized. The final analysis of OS will take place after approximately 384 and 768 death events have been observed in the 2 analysis sets,

respectively, which is estimated as 48 months after the first patient is randomized. Stopping boundaries in p-value and Z score for primary analyses of OS are shown in Table6. The boundaries will be updated according to the actual numbers of events in the interim and final analyses, using the above pre-specified alpha spending function. In the event that only OS in PD-L1+ analysis set is positive at IA, all the planed analyses in PD-L1+ analysis set will be reported. The study will continue until 768 deaths events have been observed in the ITT analysis set.

**Table 6: Stopping Boundaries (in p-value and Z score) of Primary Analysis of OS**

| Analysis Set | Analysis | Time (m) | # Events | p-value[a] (Z score) for Efficacy | p-value[a] (Z score) for Interim Futility | Approximate HR Threshold for efficacy | Cumulative Prob of Crossing Under $H_1$ |
|---|---|---|---|---|---|---|---|
| PD-L1+[b] | Interim analysis | 30 | 269 | < 0.0072 (>2.45) | > 0.5731 (< -0.18) | 0.742 | 0.47 |
| | Final analysis | 48 | 384 | < 0.0228 (> 2.00) | - | 0.815 | 0.80 |
| ITT | Interim analysis | 30 | 538 | < 0.0072 (>2.45) | > 0.5384 (< -0.1) | 0.810 | 0.56 |
| | Final analysis | 48 | 768 | < 0.0228 (> 2.00) | - | 0.866 | 0.87 |

a. 1-sided
b. PD-L1+ is 50% of the ITT

An interim analysis for OS 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.

## 9 CHANGES IN THE PLANNED ANALYSIS

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

## 10 REFERENCES

Common Terminology Criteria for Adverse Events (CTCAE). Version 4.0. United States Department of Health and Human Services, National Institutes of Health, National Cancer Institute, Washington, DC, USA, June 14, 2010.

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

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

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

# 11 APPENDIX 1 IMPUTATION RULES FOR PARTIAL DATES

## 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 > death date, then set to death date

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

## 11.2 IMPUTE PARTIAL DATES FOR ADVERSE EVENTS

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

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

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

If year of the end date is missing or end date is completely missing, do not impute. If start date of an adverse event is partially missing, impute as follows:

- If both month and day are missing and year = year of treatment start date, then set to treatment start date
- If both month and day are missing and year ≠ year of treatment start date, then set to January 01
- If day is missing and month and year = month and year of treatment start date, the set to treatment start date
- If day is missing and month and year ≠ month and year of treatment start date, the set to first of the month
- If the imputed AE start date is after AE end date (maybe imputed), then update AE start date with AE end date as final imputed AE start date. If the imputed end date >

min (death date, end of study date), then set to min (death date, end of study date).

### 11.3 IMPUTE PARTIAL DATES FOR SUBSEQUENT ANTI-CANCER SYSTEMATIC THERAPY/SURGERY/PROCEDURE

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

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

- If both month and day are missing, then set to December 31
- If only day is missing, then set to last day of the month
- If the imputed start date > min (death date, study discontinuation date, data cutoff date, start date of the next subsequent therapy), then set to min (death date, study discontinuation date, data cutoff date, start date of the next subsequent therapy)
- The imputed start date must be before or equal to the end date

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

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

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

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

- If both month and day are missing, then set to December 31
- If only day is missing, then set to the last day of the month
- For start date of prior systemic therapy for cancer, if imputed end date > randomization date – 6 months, then set to randomization date – 6 months
- For start date of prior radiotherapy/surgery, if imputed end date > randomization date, then set to randomization date - 1

If start date is partially missing, impute as follows:

- If both month and day are missing, then set to January 01
- If only day is missing, then set to the first of the month

- If the imputed start date > end date, then set to the end date

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

# 12 APPENDIX 2 RULES FOR IDENTIFYING MISSING TUMOR ASSESSMENTS

Identifying two missing tumor assessment

1) Input scheduled TA visit list for each study
    a. (eg. 6wk-12wk-18wk-27wk-36wk…; 9wk-18wk-27wk-36wk-45wk-54wk-66wk- 78wk…)
2) Identify last evaluable TA before PD or death (--LPTADT) and map it to the closest scheduled visit (--LPTADT_WK).
    a. In the event of unscheduled TA, choose the closest scheduled visit number (e.g. 6wk or 27wk) as –LPTADT_WK. It can be achieved programmatically by following the classification rule (e.g. defining thresholds) depicted in table 7 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 7 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 7 Example of scheduled tumor assessments with time window**

| Weeks | Scheduled week -1 | **Scheduled week** | Scheduled week+1 | Threshold |
|---|---|---|---|---|
| Baseline | | Baseline | | |
| Every 6 weeks for the first 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 |
| Every 9 weeks afterwards | Week 56 | **Week 57** | Week 58 | Week 61 |
| | Week 65 | Week 66 | Week 67 | Week 70 |
| | Week 74 | Week 75 | Week 76 | … |
| | | | | |