Title: A Randomized, Open-Label, Active-Controlled, Multi-Center, Phase III Clinical Study of Anti-PD-1 Antibody SHR-1210 vs. Investigator's Choice of Chemotherapy in Subjects with Locally Advanced or Metastatic Esophageal Cancer

NCT number: NCT03099382

Date: 28 Jun., 2017

PAREXEL International

Jiangsu Hengrui Pharmaceuticals Co., Ltd.

SHR-1210-III-301-ESC

A Randomized, Open-Label, Active-Controlled, Multi-Center, Phase III Clinical Study of Anti-PD-1 Antibody SHR-1210 vs. Investigator's Choice of Chemotherapy in Subjects with Locally Advanced or Metastatic Esophageal Cancer

Statistical Analysis Plan

PAREXEL Project No.: 236104

SPONSOR SIGNATURE PAGE

Approved by: _____ _____

[redacted] Date

Director of Statistics

Jiangsu Hengrui Pharmaceuticals Co., Ltd.

_____ _____

[redacted] Date

Medical Director

Jiangsu Hengrui Pharmaceuticals Co., Ltd.

PAREXEL SIGNATURE PAGE

Signatures below confirm that the review process has been completed in accordance with SOP-GDO-WW-019.

This document has been approved and signed electronically on the final page by the following:

	Signatory
Author	[Name] [redacted] Project Role: Biostatistics Lead
	Signatory
Reviewer	[Name] [redacted] Project Role: Biostatistics Reviewer
	[Name] [redacted] Project Role: Statistical Programming Lead

TABLE OF CONTENTS

ABBREVIATIONS

Abbreviation	Full Name
ADR	Adverse drug reaction
AE	Adverse event
CI	Confidence interval
CR	Complete response
eCRF	Electronic case report form
CRO	Contract research organization
D	Day
DoR	Duration of response
ECG	Electrocardiogram
EDC	Electronic data collection
ECOG	Eastern Cooperative Oncology Group
EORTC	European Organisation for Research and Treatment of Cancer
FAS	Full analysis set
HR	Hazard ratio
ICS	Informed consent set
IDMC	Independent Data Monitoring Committee
irAE	Immune-related adverse event
irRECIST	Immune-related response evaluation criteria in solid tumors
ITT	Intention-to-treat
kg	Kilogram
LVEF	Left ventricular ejection fraction
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milligram
min	Minute
mL	Milliliter
NCI-CTC	National Cancer Institute Common Terminology Criteria
ORR	Objective response rate
OS	Overall survival
P	P value
PR	Partial response
PD	Progressive disease
PFS	Progression-free survival
PPS	Per-protocol set
PT	Preferred term

Abbreviation	Full Name
QLQ	Quality of life questionnaire
RAN	Randomized set
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	Serious adverse event
SAP	Statistical analysis plan
SD	Stable disease
SOC	System organ class
SS	Safety set
TEAE	Treatment-emergent adverse event
TLF	Table, Listing, Figure

1 INTRODUCTION

This statistical analysis plan (SAP) details the methods of statistical analysis that will be used in the analysis of this study. This SAP does not include tables, listings, and figures (TLFs). TLFs will be shown in another article. Changes to the methods of statistical analysis will require amendments to the SAP.

This SAP is based on the following documents:

- Study Protocol, version 5.0 (12 Apr., 2018)
- Electronic Case Report Form (eCRF), version 7.0 (18 Feb., 2019)

2 OBJECTIVES AND ENDPOINTS

2.1 Study Objectives

2.1.1 Primary objective

To evaluate the efficacy of SHR-1210 vs. investigator's choice of chemotherapy in patients with locally advanced or metastatic esophageal cancer who are refractory to the first-line chemotherapy.

2.1.2 Secondary objectives

To evaluate the safety of SHR-1210 vs. investigator's choice of chemotherapy in patients with locally advanced or metastatic esophageal cancer who are refractory to the first-line chemotherapy.

2.1.3 Exploratory objective

To evaluate the relationship between biomarkers in tumor tissue and/or blood (such as PD-L1) and the efficacy of SHR-1210.

2.2 Study Endpoints

2.2.1 Primary endpoint

Overall survival (OS): defined as the time from the start of randomization to the death of the subject caused by any reason.

2.2.2 Secondary endpoints

Progression-free survival (PFS): defined as the time from the start of randomization to the first documented objective tumor progression or death caused by any reason, whichever occurs first.

Objective response rate (ORR): the proportion of subjects who achieve CR and PR.

Duration of response (DoR): defined as the time from the date of the first recording of tumor response to the date of the first recording of objective tumor progression or death caused by any reason.

Quality of life score: As a core scale for all cancer patients, the EORTC QLQ-C30 scale comprises 30 items, 5 functional scales (physical, role, cognitive, emotional and social), 3 symptom scales (fatigue, pain and nausea/vomiting), 1 global health status/quality of life scale and 6 single-item scales (dyspnoea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties). The EORTC QLQ-OES18 scale is mainly used for patients with esophageal cancer and contains 18 symptom items. The assessment should be completed at various time points based on the schedule of activities, starting from the first cycle until 30 days after the last dose.

2.2.3 Exploratory endpoint

To evaluate the relationship of the expression of PD-L1 and/or other biomarkers with the efficacy of SHR-1210.

Tumor tissues and/or blood obtained during the study may also be used in the future to determine tumor mutation burden, proteomics, or other analytes.

3 STUDY DESIGN

This is a randomized, open-label, active-controlled, multi-center study. A total of 438 eligible subjects with locally advanced or metastatic squamous cell carcinoma of the esophagus who are refractory to first-line chemotherapy will be randomized to investigational treatment group or control group in a 1:1 ratio and stratified by: 1. locally advanced lesion vs. distant metastasis; 2. ECOG PS (0 vs. 1).

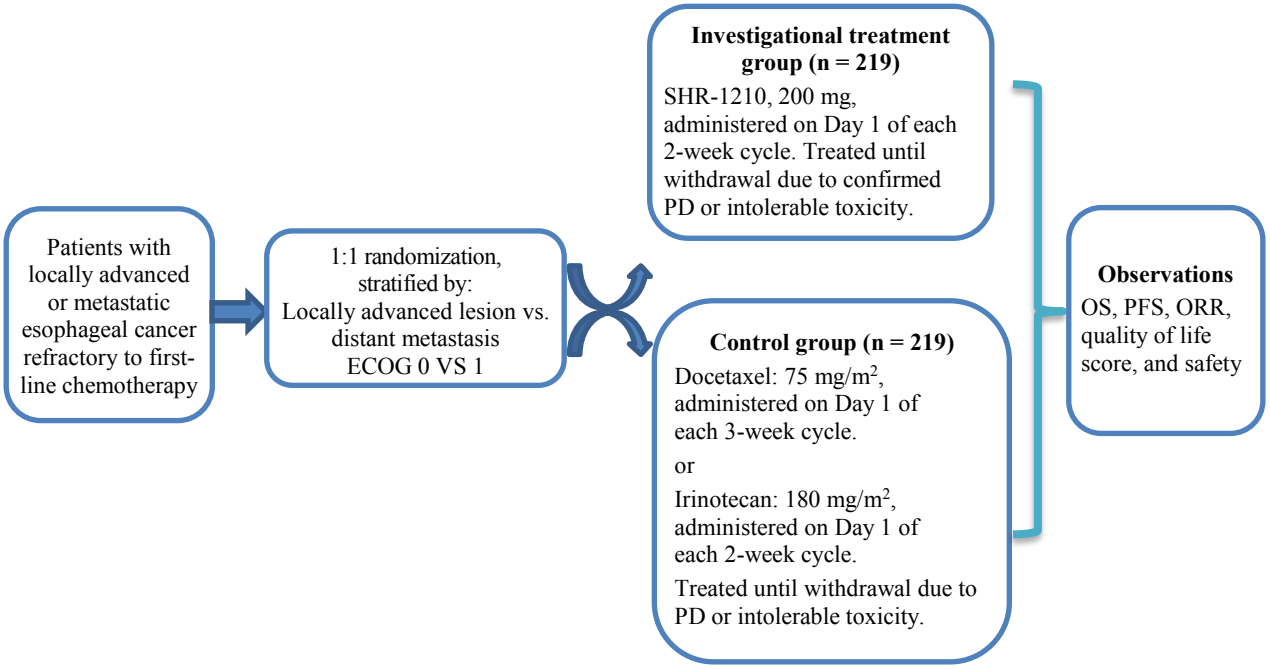
Subjects in the investigational treatment group will receive SHR-1210 (200 mg/dose, on Day 1 of each 2-week cycle) and those in the control group will receive the investigator's choice of chemotherapy, either docetaxel (75 mg/m^2, on Day 1 of each 3-week cycle) or irinotecan (180 mg/m^2, on Day 1 of each 2-week cycle). Both investigational treatment group and control group will continue treatment until the occurrence of progressive disease (PD), intolerable toxicity, voluntary withdrawal by the subject, or treatment discontinuation determined by the investigator.

Based on SHR-1210 being an immune checkpoint inhibitor and the experience of similar drugs, some subjects may experience delayed or early tumor pseudoprogression after receiving the immunotherapy. Therefore, the use of RECIST v1.1 alone may underestimate the benefits of the immunotherapy. As a result, it is suggested in this study to carry out a comprehensive imaging assessment based on both RECIST v1.1 and iRECIST, i.e., for subjects experiencing progressive disease for the first time but whose clinical symptoms are stable, the investigator may decide to

allow the subjects to continue using SHR-1210 and take the imaging assessment again after 4 weeks (±7 days); when the subsequent imaging assessment confirms progressive disease, the subjects need to discontinue SHR-1210 unless the investigator believes that the subjects can continue to gain clinical benefits from SHR-1210 and makes the decision following the discussion with the sponsor.

The study procedure is divided into three stages. The screening period is 28 days. After completing the screening examination and evaluation, the eligible subjects enter the treatment period and undergo visits by the dosing frequency. Relevant examinations and assessments must be completed before each dose. In particular, tumor imaging will be assessed once every 8 weeks (±7 days) (including the chemotherapy control group). All subjects mu st complete safety examinations and tumor assessments before the end of treatment. The subjects then enter the follow-up period. The follow-up visits of all subjects should be carried out 90 days after the last dose; the follow-up visits scheduled for 30 days after the last dose should be carried out at the study site, while those scheduled for 60 days and 90 days after the last dose via telephone; subjects ending the treatment due to reasons other than progressive disease should continue to undergo imaging assessments once every 8 weeks (±7 days) until progressive disease, initiation of new anti-cancer treatments or death; subjects ending the treatment due to progressive disease and subjects without progressive disease who complete the follow-up visits will enter the survival follow-up period, and they will be followed up once a month.

The schedule of study procedures is detailed in the protocol. The diagram of study design is shown as follows:



4 STATISTICAL METHODS

4.1 General Principles

4.1.1 Baseline

Unless otherwise stated, the "baseline" in this study is defined as the last non-missing measurement value obtained prior to the first use of study drug, including the measurements obtained on the day of and prior to the first dose. The last non-missing measurement value will be decided by the date of measurement. If there are multiple measurements on the same day, then:

- If both planned and unplanned visits are present, the value obtained from the planned visit will be considered as the last non-missing measurement.
- If there are multiple planned visits, the value obtained from the visit with the largest visit number will be considered as the last non-missing measurement.
- If there is no planned visit but there are multiple unplanned visits, the value obtained from the visit with the largest number of the unplanned visit will be considered as the last non-missing measurement.

See Section 4.1.2 for the quality of life score at baseline.

4.1.2 Visit/analysis time window

The time window will be used for the analysis summarized by visits. In the analysis carried out by visits, the statistical analysis will be performed according to the planned time points in the protocol, i.e., the time points of unplanned visits do not need to be shown.

For the analysis of quality of life scores, the analysis time window is defined as shown in the following table (calculated by the number of days of study):

Table 4.1.2-1. Analysis time window for quality of life score.

	Planned Visit Time	Analysis Time Window
Baseline	1	-7 to 1
Week 8	57	50-64
Week 16	103	96-110
Week 24	169	162-176
Week 32	225	218-232
Week 40	281	274-288
Week 48	337	330-344
…	…	…
End of Treatment (EOT)*	EOT	EOT ±7
30 Days After the Last Dose	Day of the last dose + 30	Day of the last dose + 23 to day of the last dose + 37

*End of Treatment (EOT) = day of treatment termination/withdrawal from study visit.

For the quality of life score, when the planned protocol visit is missing and there is an unplanned visit in the analysis time window of the planned visit, the unplanned visit will be used as the visit for analysis. If there are 2 or more unplanned visits in the analysis time window, the unplanned visit that is the closest to the time of planned visit will be used for visit analysis. If the time of two unplanned visits is equal to the time of the planned visit, the later unplanned visit will be used for visit analysis.

Analyses other than that of quality of life scores that should be summed up by visits (e.g. laboratory test, vital signs, ECG, physical examination, and tumor response evaluation) will be summarized by the protocol-proposed visits shown in eCRF (excluding baseline; see Section 4.1.1 for the definition of baseline) with no need to check whether the time of visit exceeds the corresponding time window specified in the protocol.

For examinations other than the quality of life score, when the visit planned in the protocol is missing, or one of the test items in the planned protocol visit is missing or its results are invalid, the unplanned visit that is the closest to the time of the planned one will be considered as the protocol-proposed visit. If the time of two unplanned visits is equally close to the time of planned visit, the later unplanned visit will be selected as the protocol visit. All visits in the protocol will be sorted in chronological order. In addition, once defined as a protocol visit, an unplanned visit will no longer be defined as other protocol visits.

For the PD-L1 expression in tumor tissues, when one subject has two or more valid tests, the result summaries will be based on the earliest one.

4.1.3 Study days

The date of administration of the first dose is defined as the start day of the study (Day 1). Then, based on the start day of the study, the number of days of study corresponding to test or event is calculated by the following formula:

- If an examination/event occurs before the start of the study, days = examination date - study start date;
- If an examination/event occurs after the start of the study, days = examination date - study start date + 1.

4.1.4 Descriptive statistics

Unless otherwise stated, all data will be summarized by treatment group and appropriate descriptive statistics based on data types. Descriptive statistics for measurement data will include number of subjects, mean, standard deviation, median, minimum, and maximum. Descriptive statistics for count data will include frequency and percentage. As for time-to-event data, the median time and its 95% confidence interval will be estimated using the Kaplan-Meier method, and the survival curve will be plotted when necessary. Meanwhile, the 95% confidence interval for the median time will be calculated using the Brookmeyer-Crowley method based on log-log transformation, and its standard error will be calculated using the Greenwood formula.

4.1.5 Number of decimal places

Unless otherwise stated, the rounding of decimal places will follow the rules below.

- The decimal places of minimum and maximum should be consistent with that of the raw data recorded on the CRF. Mean and median have one more decimal place than those of the raw data, and the standard deviation has 2 more decimal places than that of the raw data. However, there can only be at most 4 decimal places. If the value from the raw data is less than 0.0001, it will be recorded as "< 0.0001".
- The percentage will be rounded to 2 decimal places. If the frequency is 0, the percentage is not displayed.
- The P value will retain 4 decimal places. If the P value is < 0.0001, it will be expressed as "< 0.0001"; if the P value is > 0.9999, it will be reported as "> 0.9999".
- The 95% confidence interval will have one more decimal place than that of the raw data. If the raw data have no decimal place, the 95% confidence interval will retain 2 decimal places; if the raw data have 4 or more decimal places, the 95% confidence interval will retain at most 4 decimal places.

4.1.6 Analysis software

All statistical analyses will be conducted using SAS® (v9.4).

4.2 Analysis Set/Population

The following analysis sets will be involved in this study:

4.2.1 Informed consent set (ICS)

All subjects who have signed the informed consent form will be included in this analysis set.

4.2.2 Randomized set (RAN)

All subjects who have been randomized, whether or not they have received study drug, will be included in this analysis set. These subjects will be analyzed by group assigned by randomization (regardless of the actual treatment they receive).

4.2.3 Full analysis set (FAS)

All randomized subjects who have received study drug at least once will be included in FAS. The full analysis set is the primary analysis set used for the efficacy analysis of this study. In the FAS-based analysis, subjects will be analyzed by group assigned by randomization (regardless of the actual treatment they receive).

4.2.4 Per-protocol set (PPS)

PPS is a subset of FAS and defined as subjects experiencing no protocol deviations that exert major impacts on the efficacy analysis results during the study. These subjects will be analyzed by group assigned by randomization (regardless of the actual treatment they receive).

Protocol deviations with major impacts on study results may include, but are not limited to, the following three points:

- Violation of the inclusion and exclusion criteria;
- Administration of drugs prohibited by the study;
- The drugs actual used are inconsistent with the randomized assignment;

The screening and determination of subjects with protocol deviations that have major impacts on study results will be carried out before database locking, and these subjects (if any) will be removed from PPS. The PPS-based analysis will serve as a secondary/supportive analysis.

4.2.5 Safety set (SS)

All subjects who have received study drug at least once (whether they have participated in randomized grouping or not) constitute the safety set. The SS is the primary analysis population used for the safety analysis. The SS-based analysis will be performed by the treatment actually received by patients.

4.2.6 QLQ-C30 analysis set

The QLQ analysis set is a subset of FAS. It is defined as the subject population who have completed the QLQ-C30 baseline assessment and at least one post-baseline assessment. This analysis set will be used for the analysis of QLQ-C30.

4.2.7 QLQ-OES18 analysis set

The QLQ analysis set is a subset of FAS. It is defined as the subject population who have completed the QLQ-OES18 baseline assessment and at least one post-baseline assessment. This analysis set will be used for the analysis of QLQ-OES18.

4.3 Subject

4.3.1 Subject distribution

The descriptive statistics of the screening status of subjects, including number of screened subjects, number and percentage of subjects that have failed the screening, number and percentage of subjects that have undergone randomization, and number and percentage of subjects that have received treatment, will be summarized based on ICS and categorized by the reasons for screening failure (number and percentage of subjects).

The descriptive statistics of number of subjects undergoing randomization, number and percentage of subjects that have received study drug, number and percentage of subjects that have terminated the treatment, number and percentage of subjects that have withdrawn from the study, and number and percentage of subjects that have completed the study will be summarized based on RAN and the treatment groups assigned by randomization and categorized by the reasons for terminating the treatment and withdrawing from the study (number and percentage of subjects).

The descriptive statistics of number and percentage of subjects included in FAS, PPS, SS, QLQ-C30 analysis set, and QLQ-OES18 analysis set and number and percentage of subjects rejected from various analysis sets will be summarized based on RAN and treatment groups assigned by randomization and categorized by the reasons of rejection (number and percentage of subjects).

Since the control group will receive two different chemotherapy regimens, i.e., docetaxel or irinotecan, the above-mentioned summary by treatment group will be further subdivided by the chemotherapy regimen received in the control group.

A detailed listing of reasons for screening failure, randomization of subjects, withdrawal from the treatment and the reasons will be provided based on ICS.

A detailed listing of subjects included into and rejected from FAS, PPS, SS, QLQ-C30 analysis set, and QLQ-OES18 analysis set will be provided based on RAN. The listing will include rejected analysis sets and reasons for rejection at least.

4.3.2 Protocol deviations

The Hengrui project team will be responsible for recording all protocol deviations. Before the database is locked, PAREXEL will discuss with members of the Hengrui project team to determine the impact of protocol deviations on the analysis set/population. Protocol deviations can be classified as "major" or "minor". Major protocol deviation refers to protocol deviation that may have significant impacts on subjects.

A descriptive summary of major protocol deviations will be provided for all randomized subjects by treatment group and category of protocol deviations (number and percentage of subjects will be provided). In addition, a detailed listing of protocol deviations will be provided. The listing will include all major and minor protocol deviations.

4.4 Demographics and Baseline Data

The descriptive statistics of demographics and baseline data will be summarized based on FAS and by treatment and overall groups.

4.4.1 Demographics

The summary of descriptive statistics of demographics (gender, age, age stratification [< 65 years old and ≥ 65 years old] (of which, < 65 years old can be further divided into < 45, 45 to < 55, and 55 to < 65 years old), ethnicity, height [cm], body weight [kg], BMI [kg/m^2]), BMI stratification (< 18.5 kg/m^2 and ≥ 18.5 kg/m^2), smoking, and alcohol use will be provided along with detailed listings.

Age will be calculated with the number of full years between date of birth and signing date of the informed consent form. Please refer to Appendix 1 for the rules of age calculation.

Formula of BMI calculation: Baseline weight (kg)/Baseline height (m^2).

4.4.2 Tumor diagnosis

Primary tumor site, pathological grade, cumulative number of organ with metastasis, location of metastasis, and course of disease will be summarized using descriptive statistics.

Course of disease (month) is defined as the time from date of the first pathological diagnosis to date of randomization, and its formula of calculation is as follows: (Date of randomization – Date of first pathological diagnosis)/30.4375. If the "day" in the date of pathological diagnosis is missing, the course of disease will be calculated based on imputation with day 15.

A detailed listing of subjects for tumor diagnosis will be provided.

4.4.3 Medical history

Medical history will be coded by MedDRA (v20.0), and the statistics for the number and percentage of subjects under each SOC and PT will be provided. In addition, a detailed listing of subjects will be provided.

4.4.4 History of cancer treatment

History of cancer treatment mainly includes history of neoadjuvant chemotherapy, history of adjuvant chemotherapy, history of rescue therapy, history of radiotherapy, and history of tumor surgery.

For the history of cancer treatment, the following descriptive statistics will be provided:

- Number and percentage of subjects who have received neoadjuvant chemotherapy
- Number and percentage of subjects who have received adjuvant chemotherapy
- Number and percentage of subjects who have received rescue therapy
- Number and percentage of subjects who have received first-line platinum-based chemotherapy
- Number and percentage of subjects who have received radiotherapy
- Number and percentage of subjects who have received radical concurrent chemoradiotherapy. Radical concurrent chemoradiotherapy is defined as the presence of radical radiotherapy in the "History of Radiotherapy" page of eCRF and the presence of concurrent chemotherapy in the page of "Rescue Therapy", "History of Neoadjuvant Chemotherapy" or "History of Adjuvant Chemotherapy". Specifically, the concurrent chemotherapy in the "Rescue Therapy" page is defined as the treatment that shares at least one day with the radiotherapy. There is no time limit for "History of Neoadjuvant Chemotherapy" and "History of Adjuvant Chemotherapy".
- Number and percentage of subjects who have received tumor surgery

In addition, the duration between end date of the last tumor treatment and date of randomization (month) will be provided using the following formula: (Date of randomization – End date of the last tumor treatment)/30.4375.

In addition, the duration between end date of first-line chemotherapy and date of randomization (month) will be provided using the following formula: (Date of randomization – End date of first-line chemotherapy)/30.4375.

If the end date of the last tumor treatment or the end date of first-line chemotherapy is incomplete and only the "day" is missing, the aforementioned duration will be calculated based on an imputation with day 15.

A detailed listing of cancer treatment history will be provided.

4.4.5 History of allergies, medical history, and concomitant medication

The detailed listings of subjects with history of allergies, medical history, and concomitant medication will be provided separately.

4.4.6 New anti-tumor treatment

The number and percentage of subjects who have received new anti-cancer treatment after ending treatment will be summarized by the treatment type, and a detailed listing will be provided.

4.5 Treatment Compliance

The descriptive statistics of treatment compliance will be analyzed using SS.

4.5.1 SHR-1210

Dosing frequency of SHR-1210 in subjects, frequency and reasons of dose discontinuation, and frequency and reasons of treatment interruption will be summarized using descriptive statistics.

A detailed listing of SHR-1210 administration will be provided. This listing will include the reasons for dose discontinuation and treatment interruption.

4.5.2 Docetaxel

Dosing frequency of docetaxel in subjects, number and reasons of dose modifications, frequency and reasons of dose discontinuation, and frequency and reasons of treatment interruption will be summarized using descriptive statistics.

A detailed listing of docetaxel administration will be provided. The listing will include the reasons for dose modification, dose discontinuation, and treatment interruption.

4.5.3 Irinotecan

Descriptive statistics will be provided in a manner similar to that for docetaxel.

4.6 Efficacy Analysis

4.6.1 Handling of dropouts and missing data

The handling of missing data will be described separately in sections corresponding to study endpoints.

4.6.2 Multiple comparison/multiplicity

This study only involves 1 final analysis of the primary endpoint OS and multiplicity will not be involved.

The secondary analysis in this study will be only used to provide supportive conclusions for the primary analysis and no α adjustments will be made for multiplicity.

4.6.3 Subgroup analysis

The subgroup analysis is exploratory and will only be performed based on FAS. This study will perform the following subgroup analyses on OS and PFS:

First category of subgroup factors:

- Gender: male/female
- Expression of PD-L1 in tumor tissues: < 1% vs. ≥ 1%
- Cumulative organ with metastatic lesion: 1 vs. ≥ 2.
- Baseline ECOG PS: 0 vs. 1
- Clinical staging: locally advanced lesion vs. distant metastasis
- Liver metastasis: yes vs. no

Second category of subgroup factors:

- Age: < 65 years vs. ≥ 65 years
- BMI: < 18.5 kg/m^2 vs. ≥ 18.5 kg/m^2
- Smoking: Never smoke vs. Smoking vs. Quitted smoking
- Alcohol use: Never drink alcohol vs Drinking alcohol vs Quitted alcohol

To investigate the effects of subgroup factors on OS and PFS, the hazard ratios of OS and PFS of the investigational treatment group to the control group under various factors will be estimated using the Cox proportional hazards model and their 95% confidence intervals will be calculated based on the profile likelihood method. In addition, forest plots will be produced. The model will be fitted using treatment group as the only fixed effect with the randomized stratification factors being considered. If the subgroup analysis itself is one of the randomized stratification factors, the model should be equivalent to a Cox model stratified using the remaining randomized stratification factors.

4.6.4 Primary efficacy endpoint – OS

4.6.4.1 Primary analysis

The primary analysis is to evaluate the efficacy difference between investigational treatment group and control group based on the FAS. In the primary analysis, the null hypothesis that the distributions of OS are identical between the two groups will be tested using the stratified log-rank test based on randomized stratification factors (locally advanced lesion vs. distant metastasis; baseline ECOG PS (0 vs. 1).

The median survival time and its 95% confidence interval will be calculated using the Kaplan-Meier product limit method, and the survival curve will also be plotted. In addition, the number and percentage of subjects experiencing OS events and the number and percentage of censored subjects will be summarized and categorized by the reasons for censoring.

Missing data: In this study, subject survival will be tracked where possible until an OS event is observed. Inevitably, an OS event may not be observed for a subject before the end of the study or a subject drops out before an OS event is observed; in such cases, the time from the start date of randomization of this subject to the last observed survival date of the subject is taken as the subject's survival time, and this survival time is recorded as right-censored data (the actual survival time is greater than the recorded time).

OS will be calculated in months using the following formula: (Date of OS event or censored date – Date of randomization + 1)/30.4375. If the complete death date cannot be obtained, e.g. the "day" is missing, the survival time will be calculated based on day 15.

The above analysis for the primary efficacy endpoint will be repeated in FAS and PPS.

4.6.4.2 Secondary analysis

The OS secondary analysis will include:

- The 6-month, 1-year, and 2-year OS rates and their 95% confidence intervals for the two treatment groups will be estimated using the Kaplan-Meier product limit method.
- The 1-year and 2-year restricted mean survival time, its difference between treatment groups, and the corresponding 95% confidence intervals will be calculated for the two groups based on the Kaplan-Meier survival curves (regardless of the randomized stratification factors).
- The Kaplan-Meier product limit method will also be used to calculate the OS median survival time and its 95% confidence interval without the randomized stratification factors being considered, and the log-rank test method will be used to test the difference in survival distribution between groups. In addition, the differences in

survival distributions of the two groups will be compared using the Max-Combo test. The Max-Combo test will use three different Fleming-Harrington (FH) weights, namely FH (0,1), FH (1,0), and FH (1,1).

- The hazard ratio of OS of the investigational treatment group to the control group will be estimated using the Cox proportional hazards model and its 95% confidence interval will be calculated based on the profile likelihood method. The model will be fitted using the treatment group as the only fixed effect with the randomized stratification factors being considered. In addition, to explore the effects of other factors on the efficacy, the model will also include the first category of subgroup factors (except randomized stratification factors) defined in Section 4.6.3, in addition to the treatment group, as covariates to participate in the fitting, so as to test the hazard ratio of the investigational treatment group to the control group following covariate-based adjustments. The above Cox proportional hazards model will use the Efron method to process the tied data.

The above analysis for the primary efficacy endpoint will be repeated in FAS and PPS. Missing data will be recorded as right-censored data and the rules of censoring are the same as those in the primary analysis.

4.6.5 Secondary efficacy endpoints

The analysis of secondary efficacy endpoints in this study will be based on FAS only and will be carried out for all efficacy endpoints except OS (PFS, ORR, and quality of life score EORTC QLQ-C30 and EORTC QLQ-OES18).

4.6.5.1 PFS

For PFS, the inter-group differences will be compared using the log-rank test, and meanwhile, HR of PFS will be estimated and the effects of factors (identical to those for OS) other than treatment group on the differences in PFS will be investigated using a Cox proportional hazards model; both test and model are identical to those used in the OS-related analysis.

Subjects experiencing progressive disease or death (PFS events), will be censored when the following conditions occur:

- If subjects have begun new anti-cancer treatment before the PFS event, they will be censored based on the date of the last tumor assessment carried out before the start of the new anti-cancer treatment (if there is no post-baseline cancer assessment, the date of randomization will be used for censoring).

- Deaths or progressive diseases occurring after two or more planned tumor assessments are missing will be censored based on the date of the last tumor efficacy assessment carried out before deaths or progressive diseases. The missing of two or more planned tumor evaluations is defined as a situation when the interval between the event and the last tumor assessment is more than 126 days. If there is no tumor assessment after randomization, the interval from the date of baseline tumor assessment to the date of event will be calculated. If the interval exceeds 126 days, it will be recorded as censored data with censoring to the date of randomization.

If the start date of the new anti-cancer treatment is incomplete, it will be imputed with the earliest possible date.

The rules of censoring for subjects experiencing no progressive disease or death are:

- If there is no baseline tumor assessment, the date of randomization will be used for censoring.
- If subjects have not experienced a PFS event at the end date of study/analysis or before dropout, the censoring will be carried out using the date of the last tumor efficacy evaluation.
- If there is no post-baseline tumor efficacy assessment, the date of randomization will be used for censoring.

The date of progressive disease of tumor will be based on the earlier date of the following:

- The date of imaging examination showing a new lesion (if PD is based on the occurrence of a new lesion), or
- Date of the last imaging assessment of observed lesions (when progressive disease occurs based on the increased total observed lesions)

If there is no progressive disease (including PR, CR, SD, and NE), the date of tumor assessment will be the latest one recorded for target and non-target lesions (theoretically, there will be no new lesions).

4.6.5.2 ORR

Descriptive statistics will be summarized by treatment group (number and percentage of subjects). The 95% confidence interval of ORR for a single treatment group will be estimated using the Clopper-Pearson method. The 95% confidence interval of inter-group ORR difference will be estimated using the Newcombe method.

In addition, the ORR of post-treatment best overall response (BOR, the best assessment among all tumor efficacy evaluations will not include tumor efficacy evaluations carried out after progressive disease or after the start of new anti-cancer treatment) will be summarized using the same method described above.

The analytical methods identical to those described above will be used for CR, PR, SD, DCR (defined as the sum of the proportions of subjects with CR, PR, and SD in the overall tumor efficacy assessment), and PD.

In addition, the number and percentage of subjects who continue to use SHR-1210 after progressive disease and BOR of drug treatment following progressive disease will also be summarized.

Handling of missing data: When calculating CR, missing data were considered as cases not achieving CR, while the handling method for PR, SD, DCR and PD was the same as that for CR.

4.6.5.3 DoR

DoR will only be calculated for subjects who have experienced tumor response after treatment. The censoring rules for DoR are the same as those for PFS, and the end date of response of subjects must be consistent with that of progressive disease or death of subjects with PFS. For DoR, the median survival time and 6-month, 9-month, and 12-month objective response rates, as well as their 95% confidence intervals, in each treatment group will be calculated using the Kaplan-Meier product limit method, while the survival curves will be plotted. In addition, the differences between groups will be analyzed by a log-rank test using randomized stratification factors. A descriptive summary will be provided for DoR without the state of censoring being considered.

4.6.5.4 Quality of life score

4.6.5.4.1 EORTC QLQ-C30 scale

The analysis of EORTC QLQ-C30 will be performed using the QLQ-C30 analysis set.

The EORTC QLQ-C30 scale comprises 30 items, 5 functional scales (physical, role, cognitive, emotional, and social), 3 symptom scales (fatigue, pain, and nausea/vomiting), 1 global health status/quality of life, and 6 single-item scales (dyspnea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties).

For the calculation of the above scores and the handling of missing data, please refer to Appendix 3. For more information and SAS procedures, please refer to the EORTC QLQ-C30 scoring manual[1].

The descriptive statistics of the above scores at each assessment time point and their changes from baseline will be provided by treatment group along with a detailed listing.

4.6.5.4.2 EORTC QLQ-OES18 scale

The analysis of EORTC QLQ-OES18 will be performed using the QLQ-OES18 analysis set.

The EORTC QLQ-OES18 scale contains 18 symptom items. Specifically, 12 symptom items can be further categorized into 4 subscales, including dysphagia (functional scale), eating (symptom scale), reflux (symptom scale), and pain (symptom scale).

For the calculation of the scores of the above-mentioned subscales and the scores of 6 symptom items (single-item scale) and the handling of missing data, please refer to Appendix 4.

The descriptive statistics of the above scores at each assessment time point and their changes from baseline will be provided by treatment group along with a detailed listing.

4.6.6 Exploratory analysis

The exploratory analysis will be carried out based solely on FAS.

4.6.6.1 Expression level of PD-L1 in tumor tissues

For the expression of PD-L1 in tumor tissues at baseline, the relationships between the expression of PD-L1 and the OS, PFS and ORR (BOR) will be explored with cut-off points of $< 1\%$ and $\geq 1\%$, $< 5\%$ and $\geq 5\%$, and $< 10\%$ and $\geq 10\%$, respectively.

For the analysis of OS and PFS, please refer to the subgroup analysis method as shown in Section 4.6.3. The hazard ratios of OS and PFS of the investigational treatment group to the control group under various factors will be estimated using the Cox proportional hazards model and their 95% confidence intervals will be calculated based on the profile likelihood method. In addition, forest plots will be produced. Furthermore, the survival curves under various cut-off points of PD-L1 will also be plotted for OS and PFS separately based on the Kaplan-Meier product limit method.

The effects of PD-L1 expression on ORR (BOR) will be explored using a Logistic regression model. The model will include treatment group, expression of PD-L1, and interaction term between the two thereof.

4.6.6.2 Other biomarkers

This SAP will not involve the analysis of other biomarkers.

4.7 Safety Analysis

The safety analysis will be summarized by the actual treatment based on SS. Unless otherwise stated, safety will be summarized using descriptive statistics. Missing data will not be imputed.

4.7.1 Adverse events

An adverse event (AE) refers to any untoward medical condition in a clinical trial subject who receives a pharmaceutical product, and the condition does not necessarily have a causality with the treatment. AEs will be documented starting from signing of the informed consent form to 90 days after the last administration of study drugs. From 30 days after the end of treatment, only treatment-related AEs will be collected.

A treatment-emergent AE (TEAE) is defined as an AE that occurs or becomes aggravated on the same day or after the first dose of the study drug but no later than 90 days after the last use of the study drug. If the start/end date of an AE is incomplete or missing, TEAE will be judged according to the "TEAE Judgment Rules" in Appendix 2.

A treatment-related TEAE, i.e., adverse drug reaction (ADR), is defined as a TEAE whose causal relationship with the study drug is definitely related, possibly related, undeterminable, or unlikely related. If the relationship between AE and study drug is missing, the AE will be considered as an ADR.

Significant AE is defined as AE, other than serious adverse event (SAE), and significant abnormality in blood routine or other laboratory tests (significant abnormality in laboratory test will be reported as an AE) that result in the use of targeted medical measures (e.g. drug holiday, dose reduction, and symptomatic treatment).

Immune-related TEAEs of special interest include reactive capillary endothelial proliferation, immune-related pneumonitis, immune-related colitis, immune-related hepatitis, immune-related myocarditis, immune-related nephritis and renal dysfunction, immune-related endocrine diseases, immune-related cutaneous adverse drug reactions, and infusion reactions.

For the summary statistics on the incidence of AE, each subject will be counted at most once under each SOC and PT. One AE that occurs multiple times in one subject will be counted based on its most severe level for the summary statistics of severity. When the severity of AEs is calculated, one AE that occurs multiple times in one subject will be counted based on its most severe level. If the severity of the AE that occurs once or multiple times in one subject is missing, no imputation will be made and summary will be carried out with "missing" as the category.

PT and SOC of AEs will be coded using MedDRA (v20.0), and the dictionary version will be updated once at the end of the project without in-project version alteration. The severity of AEs will be graded by NCI CTCAE (v4.03).

First, the number and percentage of subjects who have experienced at least one of the following TEAEs throughout the study will be summarized:

- Any TEAE
- Treatment-related TEAE (ADR)
- Grade ≥ 3 TEAE
- Grade ≥ 3 ADR
- TEAE leading to treatment interruption/discontinuation
- ADR leading to treatment interruption/discontinuation
- TEAE leading to death
- ADR leading to death
- Any TESAE (including death)
- Treatment-related TESAE (including death)
- TESAE leading to treatment interruption/discontinuation
- TESAE leading to death
- Treatment-related TESAE leading to death
- TEAE of special interest
- Investigator-judged immune-related TEAE
- Investigator-judged immune-related TEAE leading to death
- Significant TEAE
- Treatment-related significant AE (significant ADR)
- Event with the highest CTC grade: grade 1-5

Second, the number and percentage of subjects under each category of the following AEs will be summarized:

- TEAE (at least one) by SOC and PT
- ADR (at least one) by SOC and PT

- Grade ≥ 3 TEAE (at least one) by SOC and PT
- Grade ≥ 3 ADR (at least one) by SOC and PT
- TEAE by SOC and PT (at least one, categorized by the most severe level)
- Treatment-related TEAE (at least one, categorized by the most severe level)
- TEAE leading to treatment suspension by SOC and PT
- TEAE leading to treatment termination by SOC and PT
- ADR leading to treatment suspension by SOC and PT
- ADR leading to treatment termination by SOC and PT
- TEAE leading to death by SOC and PT
- ADR leading to death by SOC and PT
- TESAE (at least one) by SOC and PT
- Treatment-related TESAE (at least one) by SOC and PT
- TESAE leading to treatment suspension by SOC and PT
- TESAE leading to treatment termination by SOC and PT
- TESAE leading to death by SOC and PT
- Treatment-related TESAE leading to death by SOC and PT
- TEAE of special interest by SOC and PT
- TEAE of special interest leading to treatment suspension by SOC and PT
- TEAE of special interest leading to treatment termination by SOC and PT
- TEAE of special interest by SOC and PT (including outcome and corrective treatment)
- Investigator-judged immune-related TEAE by SOC and PT
- Investigator-judged immune-related TEAE leading to treatment suspension by SOC and PT
- Investigator-judged immune-related TEAE leading to treatment termination by SOC and PT
- Investigator-judged immune-related TEAE leading to death by immune-related TEAE and PT

- Investigator-judged immune-related TEAE by SOC and PT (including outcome and corrective treatment)
- Significant TEAE by SOC and PT
- Significant ADR by SOC and PT

Then, the following AEs will be summarized by SOC and PT, and the number and percentage of subjects under each PT will be provided:

- TEAE by SOC and PT with an incidence of ≥ 5%
- ADR by SOC and PT with an incidence of ≥ 2%
- TEAE by SOC and PT with incidence of ≥ 10% in the treatment group and inter-group incidence difference of ≥ 5%
- ADR by SOC and PT with incidence of ≥ 10% in the treatment group and inter-group incidence difference of ≥ 5%
- Grade ≥ 3 TEAE by SOC and PT with incidence of ≥ 10% in the treatment group and inter-group incidence difference of ≥ 2%
- Grade ≥ 3 ADR by SOC and PT with incidence of ≥ 10% in the treatment group and inter-group incidence difference of ≥ 2%

Meanwhile, the following AEs will be summarized based on immune-related TEAE of special interest, and the number and percentage of these AEs will be provided.

- Immune-related TEAE of special interest (at least one) by immune-related TEAE and PT, as well as the most severe CTCAE grade
- Immune-related TEAE of special interest leading to treatment interruption/discontinuation by immune-related TEAE
- Immune-related TEAE of special interest by immune-related TEAE (including the alleviated number and percentage)

In addition, for immune-related TEAE of special interest (by immune-related TEAE), time to the occurrence of AE, time to the occurrence of grade ≥ 3 AE, duration of AE, duration of grade ≥ 3 AE, response time of AE, response time of grade ≥ 3 AE, incidence of AE calculated by exposed person-year, and incidence of grade ≥ 3 AE calculated by exposed person-year will be summarized. The incidences of TESAE and treatment-related TESAE will also be calculated by exposed person-year and summarized by SOC and PT.

The time (day) to the occurrence of AE is defined as the time from the date of the first study drug administration to the date of the first occurrence of the AE. If one subject has experienced multiple times of one AE, the date of the first occurrence of the AE will be used for calculation regardless of the AE grading. The formula is as follows: Start date of AE – Date of the first dose of study drug + 1.

The time (day) to the occurrence of grade ≥ 3 AE is defined as the time from the date of the first study drug administration to the date of the first occurrence of the grade ≥ 3 AE. The calculation method is similar to the time to the occurrence of AE.

The response time of AE (day) is defined as the date from the start to the end of the AE, and the calculation formula is as follow: End date of AE – Start date of AE + 1. If one subject has experienced multiple times of one AE, the cumulative time of the AE will be calculated, but the time overlapping of multiple times of the AE should be excluded in the calculation. If an AE is not relieved by the cut-off time point of analysis or the end of the study (the response of the AE will be judged by the last outcome information of the AE, and the response include AE outcome of recovered/resolved, recovered/resolved with sequelae, and alleviated), the end date of the AE will be used for censoring. If the end date of the AE is missing, the date of the last follow-up visit or the date of death (whichever is earlier) will be used for censoring. If one AE category (such as the category of immune-related AE) contains more than one AE (by PT) and one of these AEs is not relieved, this AE category will be judged as not relieved.

The duration of AE is defined similarly to the response time of AE, with the difference that the duration of AE does not need to consider the state of censoring.

Time to the occurrence of AE, time to the occurrence of grade ≥ 3 AE, response time of AE, and response time of grade ≥ 3 AE will only be calculated for the population who have experienced AE, and the corresponding median time and 95% confidence interval will be estimated by the Kaplan-Meier product limit method. For duration of AE and duration of grade ≥ 3 AE, a descriptive summary will be provided as measurement data.

The incidence of AE calculated by exposed person-year is the ratio of the number of subjects experiencing this AE to the total exposed time by person-year. The total exposed time by person-year is defined as the sum of time from the first dose of the study drug to 90 days after the end of study drug treatment for all subjects. For subjects who have withdrawn from the study, died within 90 days after the end of study drug treatment, or been lost to follow-up, the time between the date of the first drug administration to the date of the last follow-up visit or death, whichever is earlier, will be calculated. The total exposed time by person-year will be calculated in years, and the number of days will be converted into the number of years by dividing 365.25.

For immune-related TEAE of special interest (by immune-related TEAE), the descriptive summary of the number and percentage of subjects receiving high dose of corticosteroids (defined as at least one dose or a single daily dose of ≥ 40 mg of the equivalent dose of prednisone), the starting dose of corticosteroids, and the duration of dosing will also be provided. The dosage of different hormonal drugs will be converted into the equivalent dose of prednisone according to the following formula: 0.75 mg of dexamethasone = 5 mg of prednisone = 5 mg of prednisolone = 4 mg of methylprednisolone = 20 mg of hydrocortisone = 0.8 mg of betamethasone.

If one subject experiences one AE and receives two or more hormones during the presence of the AE, the starting dose of the hormone used first will be involved in the statistics of starting doses (i.e., the starting dose(s) of subsequent hormone therapy/therapies will not be involved in the statistics of starting doses). Then, the cumulative duration of the use of multiple hormones (time overlapping of hormone therapies should be excluded in the calculation) will be summarized as the duration of dosing. If one AE category (such as the category of immune-related AE) contains more than one AE (by PT), this AE category will be treated as one AE when being summarized. When the starting dose is summarized, if the dosing frequency of hormones is more than once a day, it will be converted into a single daily dose before being summarized. When the duration of dosing is summarized, if the end date of hormone drug treatment is unknown, it will be imputed with the start date of the last dose of the hormone drug treatment.

The corrective treatment will also be summarized for immune-related endocrine diseases. Corrective treatment, incidence of bleeding, and incidence of infection will also be summarized for reactive capillary endothelial proliferation.

A detailed listing of AE, SAE, AE leading to treatment suspense/termination, and death will be provided.

4.7.2 Laboratory measurements

For blood routine, blood biochemistry, urinalysis, and thyroid functions, a shift table will be used to descriptively summarize the number and percentage of subjects showing most significant changes from baseline after the first administration. Among them, the statistics of the most severe grading will be calculated for indicators whose toxicity grading can be carried out by CTCAE. The statistics of the most severe clinical significance (grading of clinical severity: abnormal with clinical significance > abnormal with no clinical significance > normal) will also be provided for the above indicators. In addition, the most severe grade 3/4 conditions occurring after the first administration will be summarized for the CTCAE grading.

In addition, for the judgment of clinical significance of blood routine, blood biochemistry, urinalysis, and thyroid functions, a shift table by treatment group and test/analysis time points will be used to descriptively summarize the number and percentage of subjects showing changes

from baseline. For quantitative test/analysis indicators such as blood routine, blood biochemistry, and thyroid functions, the measurements of various indicators and their change from baseline will be descriptively summarized by treatment group and test/analysis time point.

The judgment of clinical significance of test indicators such as blood routine, blood biochemistry, urinalysis, thyroid function, coagulation function, fecal occult blood, and virology will be provided with a detailed listing of various laboratory test items. The listing will include at least name of visit, date of sampling, number of days of study, measurement and unit (if applicable), judgment of clinical significance (if applicable), and CTCAE severity grading (if any).

4.7.3 ECOG (performance status)

ECOG PS will be treated as a categorical variable, and the number and percentage of subjects with various ECOG PS will be summarized by treatment group and assessment time point. In addition, a detailed listing of ECOG will also be provided. The listing will include at least name of visit, date of scoring, study day, and ECOG PS.

4.7.4 Vital signs

The vital signs to be collected in this study will include body temperature, heart rate, respiration, diastolic blood pressure, and systolic blood pressure. The number and incidence of subjects with clinically significant abnormalities after the first administration will be descriptively summarized by treatment group (refer to the following criteria for determining clinically significant abnormalities in vital signs), along with the results obtained at various time points and their changes from baseline.

In addition, a detailed listing of vital signs will also be provided. The listing will include at least the name of the visit, the date of the examination, the study day, the results of various examination items of vital signs, and the signs of clinically significant abnormalities.

The criteria for determining clinically significant abnormalities of vital signs are as follows:

Test Item	Unit	Low	High
Body Temperature	℃	≤ 35 ˚C and change from baseline ≤ -1.1 ˚C	≥ 38.3 ˚C and change from baseline ≥ 1.1 ˚C
Heart Rate	bpm	< 40 bpm	> 120 bpm
Respiration	resp/min	< 12 times/min	> 24 times/min
Systolic Blood Pressure	mmHg	< 90 mmHg or change from baseline ≤ -30 mmHg	change from baseline ≥ 30 mmHg
Diastolic Blood Pressure	mmHg	< 50 mmHg or change from baseline ≤ -20 mmHg	change from baseline ≥ 20 mmHg

4.7.5 12-lead ECG

The descriptive summary of the number and percentage of subjects based on the most severe cases after the first administration will be provided by clinical significance judgment, i.e., normal, abnormal with no clinical significance, and abnormal with clinical significance (grading of clinical severity: abnormal with clinical significance > abnormal with no clinical significance > normal). The clinical significance judgment will also be descriptively summarized by test time points.

For PR interval and QTcF, the incidence of each category defined in the table below after the first administration will be summarized. In addition, the incidence will be summarized by test time points.

Classification criteria of PR and QTcF

Test Item	Unit		Criteria for Abnormality
PR	ms	Absolute value	max. ≥ 300
		Change from baseline	Baseline > 200 and change from baseline ≥ 25%, or, baseline ≤ 200 and change from baseline ≥ 50%
QTcF	ms	Absolute value	450 ≤ max. < 480 480 ≤ max. < 500 max. ≥ 500
		Change from baseline	30 ≤ max. < 60 max. ≥ 60

For ECG parameters such as heart rate, PR interval, QT interval, and QTcF, the results obtained at various time points and their changes from baseline will be descriptively summarized.

A detailed listing will be provided for the results of ECG.

4.7.6 Echocardiography

A detailed listing of echocardiogram will be provided.

4.7.7 Physical examination

A shift table by treatment group and time points will be used to descriptively summarize the number and percentage of subjects showing changes from baseline in various investigations (normal, abnormal), and a detailed listing will be provided.

4.8 Determination of Sample Size

This is parallel study with its primary endpoint being OS. Subjects will be randomized in a 1:1 ratio. Assuming that the median OS of the control group is 7 months and the median OS of the treatment group is 9.5 months (the treatment group will be extended by 2.5 months compared with that of the control group), the inter-group OS distribution will be compared using a

stratified log-rank test at an overall significance level $\alpha = 0.025$ (one-sided); when the duration of enrollment is assumed to be 18 months and the entire study lasts for 36 months, at least 365 OS events need to be collected to obtain a power of 80% according to the calculation by East (v6.3). Furthermore, assuming a drop-out rate of 20%, a total of 438 subjects need to be enrolled in this study (219 in the investigational treatment group and 219 in the control group).

4.9 Updates for Study Implementation or Planned Analysis

Since the immunogenicity data are not collected during the study implementation, the immunogenicity analysis mentioned in Section 11.4.4 of the protocol will no longer available.

A DoR analysis is added for the secondary endpoints.

5 REFERENCES

[1] EORTC. The EORTC QLQ-C30 Scoring Manual. 3rd edition. 2001.

[2] EORTC. Scoring Procedure for the EORTC-QLQ-OES18.

6 APPENDICES

6.1 Appendix 1

Rules of Age Calculation:

Date of Informed Consent	Date of Birth	Algorithm
Complete	Complete	When the "month" and "day" in the date of the informed consent are equal to or later than those in the date of birth, age = number of years in the date of informed consent – number of years in the date of birth. When the "month" and "day" in the date of the informed consent are earlier than those in the date of birth, age = number of years in the date of informed consent – number of years in the date of birth – 1.
	Incomplete	When the "day" in the date of birth is missing, the calculation will be carried out using the 15th day of the corresponding month. The rules of calculation are the same as those applicable to the calculation of complete date.

6.2 Appendix 2

Rules of Imputing Incomplete Date:

Date	Earliest Possible Date	Latest Possible Date
Missing "Day"	The first day of the month	The last day of the month
Missing "Month" and "Day"	1 Jan.	31 Dec.

TEAE Judgment Rules

Start/Aggravation Date	End Date	Rule
Complete	Complete	When "start date" ≥ "date of first study drug administration" and "start date" ≤ "date of last study drug administration" + 90, it is judged as a TEAE.
	Incomplete	TEAE is judged based on "start date", and "end date" is imputed with the latest possible date.
	Missing	TEAE is judged based on the "start date".
Incomplete, but it can be determined from the known part of the date that the AE is impossible to occur on the same day of or after the first study drug administration	Complete	It's not a TEAE. "Start date" is imputed with the earliest possible date.
	Incomplete	It's not a TEAE. "Start date" is imputed with the earliest possible date. "End date" is imputed with the latest possible date.
	Missing	It's not a TEAE. "Start date" is imputed with the earliest possible date.
Incomplete, but AE may occur on the same day of or after the first study drug administration	Complete	When "end date" ≥ "date of first study drug administration", it is judged as a TEAE. The earliest possible date is used as "start date". When the earliest possible date is earlier than "date of first study drug administration", "date of first study drug administration" will be used for imputation. When "end date" < "date of first study drug administration", it is not judged as a TEAE; "start date" is imputed with the earliest possible date.
	Incomplete	It is judged as a TEAE. "Start date" is imputed with the earliest possible date; when the earliest possible date is before "date of first study drug administration", "date of first study drug administration" will be used for imputation; "end date" is imputed with the latest possible date.
	Missing	It is judged as a TEAE. "Start date" is imputed with the earliest possible date; when the earliest possible date is before "date of first study drug administration", "date of first study drug imputation" will be used for imputation.
Missing	Complete	When "end date" ≥ "date of first study drug administration", it is judged as a TEAE.
	Incomplete	When "end date" is incomplete, the latest possible date will be used for imputation; when the latest possible date ≥ "date of first study drug administration", it is judged as a TEAE.
	Missing	It is judged as a TEAE.

6.3 Appendix 3

EORTC QLQ-C30 (v3) Scoring Guideline

	Scale	Number of Items	Item Range	eCRF Item Numbers	Functional Scales
Global Health Status	QL2	2	6	29, 30	
Functional Scales					
	PF2	5	3	1 to 5	F
	RF2	2	3	6, 7	F
	EF	4	3	21 to 24	F
	CF	2	3	20, 25	F
	SF	2	3	26, 27	F
Symptom Scales					
	FA	3	3	10, 12, 18	
	NV	2	3	14, 15	
	PA	2	3	9, 19	
Single-Item Scale					
	DY	1	3	8	
	SL	1	3	11	
	AP	1	3	13	
	CO	1	3	16	
	DI	1	3	17	
	FI	1	3	28	

For all scales, the calculation of the scores is divided into the following steps:

- Step 1: Calculate the RawScore = RS = $I_1 + I_2 + \ldots + I_n) / n$ of each scale, where n is the number of items answered in each scale, and I represents the score of each item in this scale.
- Step 2: For Functional scales, the final score is calculated by the following formula: Score = $(1 – (RS – 1)/range) \times 100$, where range is the value range of RS, equivalent to the value range of a single item in the functional scales. For Global health status, Symptom scales, and 6 Single-item scales, the final score is calculated using the following formula: Score = $((RS – 1)/range) \times 100$.

Handling of missing data: For the global health status, functional scale, and symptom scale, if at least half of the items from the scale have been answered, the score is calculated using the above steps; otherwise, the score of this scale is considered missing.

6.4 Appendix 4

EORTC QLQ-OES18 Scoring Guideline

	Scale	Number of Items	Item Range	eCRF Item Numbers
Functional Scales				
	OESDYS	3	3	31 to 33
Symptom Scales				
	OESEAT	4	3	36 to 39
	OESRFX	2	3	44, 45
	OESPA	3	3	46 to 48
Single-Item Scale				
	OESSV	1	3	34
	OESCH	1	3	35
	OESDM	1	3	40
	OESTA	1	3	41
	OESCO	1	3	42
	OESSP	1	3	43

The calculation methods and handling of missing data of Functional scales, Symptom scales, and Single-item scales are the same as those shown in Appendix 3.