

Statistical Analysis Plan

Version 1.0

Name of study drug	TAS-118, Oxaliplatin
Study title	An open-label randomized multi-center phase III study of TAS-118 plus Oxaliplatin versus S-1 plus Cisplatin as first-line therapy in patients with advanced gastric cancer
Protocol number	[REDACTED]
Date of preparation or revision	February 4, 2019
Author	[REDACTED]

Date and signature

[REDACTED] [REDACTED] [REDACTED]	Month Day, Year	
[REDACTED] [REDACTED] [REDACTED] [REDACTED]	Month Day, Year	

Preparation and revision history

Date	Author	Reason for preparation or revision
February 4, 2019	[REDACTED]	Prepared the first version

TABLE OF CONTENTS

1. Scope and Objective	7
1.1. Positioning of Statistical Analysis Plan.....	7
1.2. Study Objective	7
1.3. Changes from Protocol	7
2. Main Statistical Considerations in the Study Design and Planned Sample Size	7
2.1. Planned Sample Size	7
2.1.1. Planned Sample Size	7
2.1.2. Rationale for Sample Size Determination Before Study Initiation.....	7
2.1.3. Change in the Required Number of Events in the Analysis for Primary Evaluation.....	9
2.2. Subject Registration and Randomization Method	10
2.2.1. Randomization Method	10
3. Data to Be Used in Statistical Analysis and Analysis Populations	11
3.1. Materials and Data Other Than Case Report Forms.....	11
3.2. Analysis Populations	12
4. Timing of Analyses and Significance Level	13
5. Endpoints and Common Considerations for Statistical Analysis	14
5.1. Endpoints.....	14
5.1.1. Efficacy Endpoints	14
5.1.1.1 OS	14
5.1.1.2 PFS	14
5.1.1.3 TTF	15
5.1.1.4 Best Overall Response Assessment.....	15
5.1.1.4.1 ORR.....	15
5.1.1.4.2 DCR.....	16
5.1.1.4.3 Duration of Overall Response	16
5.1.1.4.4 Duration of Complete Response.....	16
5.1.1.4.5 Tumor Shrinkage Rate.....	16
5.1.2. Safety Endpoints.....	16
5.1.2.1 Adverse Events and Adverse Drug Reactions	16
5.1.3. Compliance Endpoints.....	17
5.1.3.1 Total Dose.....	17
5.1.3.2 Dose Intensity	17
5.2. Common Considerations for Statistical Analysis	18
5.2.1. Data Handling.....	18
5.2.1.1 Handling of Longitudinal Data.....	18
5.2.1.1.1 Calculation of Number of Days and Period.....	18
5.2.1.1.2 Calculation of Number of Years, Months, and Weeks	18
5.2.1.1.3 Acceptable Time Window	19
5.2.1.2 Handling of Missing Data and Abnormal Values.....	19
5.2.1.3 Handling of Treatment Arms.....	20
5.2.2. Patient Baseline Characteristics.....	20
5.2.3. Statistical Methods	22
5.2.3.1 Summary Statistics for Continuous Data.....	22
5.2.3.2 Proportion	22
5.2.3.3 Test Methods	22

5.2.3.4	Model Analysis.....	24
5.2.3.5	Estimation Methods.....	24
5.2.4.	Sample Code.....	25
5.2.4.1	Stratified Log-Rank Test	25
5.2.4.2	Log-Rank Test	25
5.2.4.3	Cox Proportional Hazards Model	25
5.2.4.4	Cox Proportional Hazards Model (by Baseline Characteristic Factor)	25
5.2.4.5	Restricted Mean Survival Time (RMST)	26
5.2.4.6	95% Confidence Interval of Between-Group Difference in MST by Bootstrap Method.....	27
5.2.4.7	Stratified Cox Proportional Hazards Model	27
5.2.4.8	Comparison of Proportions Between Treatment Arms.....	28
5.2.4.9	Confidence Interval of Proportion	28
5.2.4.10	Wilcoxon Test (Two-Sample)	28
5.2.4.11	Confidence Interval of Median.....	28
5.2.4.12	Log-Rank Test Using Recalculated α	28
5.2.4.13	Cochran–Mantel–Haenszel Test.....	31
5.2.5.	Software to Be Used	31
5.2.6.	Dictionaries to Be Used.....	31
6.	Primary Endpoint Analyses.....	33
6.1.	Primary Analysis for Primary Objective	33
6.2.	Sensitivity Analyses of Primary Analysis	33
6.3.	Secondary Analyses of Primary Endpoint.....	33
7.	Secondary Endpoint Analysis.....	35
7.1.	Efficacy Analyses	35
7.1.1.	Antitumor Activity	35
7.1.1.1	Best Overall Response.....	35
7.1.1.2	ORR.....	35
7.1.1.3	DCR.....	35
7.1.1.4	Time to Response	35
7.1.1.5	Duration of Overall Response	35
7.1.2.	PFS	35
7.1.3.	TTF	36
7.1.4.	Tumor Shrinkage Rate.....	36
7.2.	Safety Analysis.....	36
7.2.1.	Adverse Events and Adverse Drug Reactions	36
7.2.2.	Deaths	37
7.3.	Composition of Analysis Populations	37
7.4.	Deviation	37
7.5.	Analyses of Patient Baseline Characteristics.....	37
7.6.	Compliance Analyses	37
7.6.1.	Dose Intensity	37
7.6.2.	Compliance.....	37
7.6.3.	Discontinuation.....	37
7.6.4.	Cumulative Dose/Duration of Treatment	37
7.6.5.	Treatment Completion.....	38
7.6.6.	Treatment Cycle	38
7.6.7.	Proportion of Outpatient Visits.....	38
7.7.	Concomitant Medications.....	38

7.8.	Subsequent Therapy	39
7.9.	Laboratory Values	39
7.10.	Clinical Findings	40
7.11.	Tumor Markers	40
7.12.	Vital Signs and ECOG PS	40
8.	Review and Revision of Statistical Analysis Plan	40
9.	Development of Output Plan	40
10.	References	40

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

The abbreviations and terms used in this statistical analysis plan are defined below.

Abbreviation (term)	Explanation (definition)
BSA	Body surface area
CDDP	(<i>SP-4-2</i>)-Diamminedichloroplatinum, cisplatin
CR	Complete response (response associated with antitumor activity)
CRP	C-reactive protein
DCR	Disease control rate
ECOG	Eastern Cooperative Oncology Group
FAS	Full analysis set
HR	Hazard ratio
IDMC	Independent data monitoring committee
IWRS	Interactive web response system
L-OHP	(<i>SP-4-2</i>)-[(1 <i>R</i> , 2 <i>R</i>)-Cyclohexane-1, 2-diamine- κ N, κ N'] [ethanedioato(2-)- κ O ¹ , κ O ²]platinum: oxaliplatin (JAN)
MedDRA	Medical Dictionary for Regulatory Activities
MST	Median survival time
NE	Not evaluable (response associated with antitumor activity)
ORR	Overall response rate
OS	Overall survival
PD	Progressive disease (response associated with antitumor activity)
PFS	Progression-free survival
PR	Partial response
PS	Performance status
RECIST	Response evaluation criteria in solid tumors ¹⁾
RMST	Restricted mean survival time
SD	Stable disease (response associated with antitumor activity)
SOC	System organ class
TTF	Time to treatment failure

1. Scope and Objective

1.1. Positioning of Statistical Analysis Plan

This statistical analysis plan provides the details of the statistical analyses to be conducted in the primary and final evaluations in “An open-label randomized multi-center phase III study of TAS-118 plus Oxaliplatin versus S-1 plus Cisplatin as first-line therapy in patients with advanced gastric cancer” [REDACTED].

1.2. Study Objective

To compare the following endpoints for TAS-118/oxaliplatin (L-OHP) (study arm) with S-1/cisplatin (CDDP) (control arm) in patients with advanced gastric cancer:

Primary

- Overall survival (OS)

Secondary

- Progression-free survival (PFS)
- Time to treatment failure (TTF)
- Overall response rate (ORR)
- Disease control rate (DCR)
- Safety

1.3. Changes from Protocol

The full analysis set (FAS) is defined as “patients in as-treated (AT) population who have gastric cancer at randomization” in the protocol version 3.0, but the definition is changed to “patients in AT population who have advanced gastric cancer at randomization” in this statistical analysis plan. Since this study involves patients with advanced gastric cancer as specified in the primary objective, the protocol has not been revised, and the detailed definition of FAS is specified in the statistical analysis plan as determined in the discussion with medical advisor and the statistical advisor.

2. Main Statistical Considerations in the Study Design and Planned Sample Size

2.1. Planned Sample Size

2.1.1. Planned Sample Size

Target number of patients at study initiation: 686 subjects (TAS-118/L-OHP arm: 343 subjects, S-1/CDDP arm: 343 subjects); target number of patients included in the FAS: 650 subjects (TAS-118/L-OHP arm: 325 subjects, S-1/CDDP arm: 325 subjects)

2.1.2. Rationale for Sample Size Determination Before Study Initiation

In a randomized phase II study of S-1/leucovorin (LV) therapy, S-1/LV/L-OHP (SOL) therapy, and S-1/CDDP therapy in patients with unresectable, advanced/recurrent gastric cancer ([REDACTED] [REDACTED]) the median survival time (MST) was 12.6 months in the S-1/CDDP arm and 18.4 months in the SOL arm. In the SOS trial, which was conducted in Japan and Korea,⁴⁾ the MST was 13.9 months in the S-1/CDDP (5-week interval) arm (15.5 months in Japanese subjects). Taking into account the ratio of the sample size between Japan and Korea and the target population in the present study, the MST is assumed to be 14.5 months in the S-1/CDDP arm and 19.33 months in the TAS-118/L-OHP arm (hazard ratio of TAS-118/L-OHP versus S-1/CDDP, 0.75).

With an overall significance level of 5% (two-sided) in the entire study, the Lan and DeMets alpha spending function (O'Brien-Fleming type)⁵⁾ was used to split the significance level for one benefit

evaluation in an interim analysis. Assuming that test statistics in the interim analysis and the primary analysis for the primary evaluation would follow a bivariate normal distribution, the boundaries $z_{\alpha 1}$ and $z_{\alpha 2}$ that meet the formula below are 2.9627 and 1.9687, respectively, and the corresponding nominal significance levels are 0.305% and 4.899% (two-sided), respectively, in the interim analysis and the primary analysis for the primary evaluation. The information fraction is assumed to be 0.5 because the interim analysis will be conducted when half the specified number of events have been observed.

$$P(Z_1 \geq z_{\alpha 1}) = 2 \times \left\{ 1 - \Phi \left(\frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{t}} \right) \right\} = 0.001525$$

$$P(Z_1 \geq z_{\alpha 1}, Z_2 \geq z_{\alpha 2}) = 0.025$$

$\alpha 1$: one- sided significance level in an interim analysis,

$\alpha 2$: one- sided significance level in the primary analysis for the primary evaluation,

t : information fraction, α : overall one- sided significance level

On the condition that the survival time is exponentially distributed in both arms, with a two-sided significance level of 4.899% as calculated above and 85% power, 437 events will be required to test whether TAS-118/L-OHP is superior to S-1/CDDP with respect to survival time in the primary evaluation, as calculated using the Schoenfeld method below:⁶⁾

$$\text{Required number of events} = \frac{\left(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta} \right)^2}{r_1 \times r_2 \times (\log(\Lambda))^2}$$

α : two- sided significance level, $1-\beta$: power, Λ : hazard ratio,

r_1 : randomization ratio in Arm A, r_2 : randomization ration in Arm B

Assuming that patients would be uniformly accumulated for an accrual period of 17 months and be followed up for 19 months, the target sample size is estimated to be 650 patients according to the formula below.⁷⁾ In this case, approximately 38 patients need to be enrolled per month; however, the accrual period is set at 19 months because accrual may be slower than required and approximately 5% of patients may be additionally excluded from the FAS, and therefore, the total sample size is calculated to be 686 patients in the overall two groups.

$$\text{Target sample size} = \text{Required number of events} / \sum_i r_i \left(1 - \frac{e^{-\lambda_i Y} (1 - e^{-\lambda_i X})}{\lambda_i X} \right)$$

i : arm, r_i : randomization ratio in Arm i , λ_i : hazard in Arm i ,

X : accrual period, Y : follow-up period

Subjects can be additionally enrolled when death events occur less frequently than expected. If the number of subjects to be enrolled is increased above 686 subjects, the sponsor will determine the final planned sample size in the discussion with the statistical advisor.

Null and alternative hypotheses in statistical tests are as follows:

- Null hypothesis H_0 : $\lambda_T/\lambda_C = 1$
- Alternative hypothesis H_1 : $\lambda_T/\lambda_C \neq 1$
- λ_C : Hazard with S-1/CDDP
- λ_T : Hazard with TAS-118/L-OHP

A list of powers provided when expected MST is changed in each arm is shown in Table 2-1.

Table 2-1 Changes in power (%) in the final analysis when expected MST is changed

MST (months) in TAS-118/L-OHP arm	MST (months) in S-1/CDDP arm						
	12	13	14	14.5	15	16	17
17	95.3	79.8	52.4	38	25.4	9.1	2.4
18	98.8	92.4	74.5	61.5	47.5	23	8.5
19	99.8	97.7	88.9	80.4	69.2	43.2	21
19.33	99.9	98.5	92	85	75.2	50.3	26.6
20	100	99.4	96.1	91.8	85	64.2	39.4
21	100	99.9	98.8	97.1	93.9	80.9	59.5
22	100	100	99.7	99.2	97.9	91.3	76.6

The powers were calculated with a total number of events of 437 and a two-sided significance level of 4.899%.

The power selected in this study is underlined.

The overall power (a probability that the test is significant under the alternative hypothesis either in the interim analysis or the primary evaluation) is 85.1% in this study, including the interim analysis. This power was calculated using logarithmic transformation and normal approximation of the expected hazard ratio according to the following formulas:

$$\text{Total power} = 1 - \Pr\left(Z_1 > z_{\frac{\alpha_1}{2}}, Z_2 > z_{\frac{\alpha_2}{2}}\right)$$

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \log(\Lambda) \times \sqrt{\frac{N_1}{4}} \\ \log(\Lambda) \times \sqrt{\frac{N_2}{4}} \end{pmatrix}, \begin{pmatrix} 1 & \sqrt{\frac{N_1}{N_2}} \\ \sqrt{\frac{N_1}{N_2}} & 1 \end{pmatrix}\right),$$

α_1 : two-sided significance level in an interim analysis,

α_2 : two-sided significance level in the primary analysis for the primary evaluation,

N_1 : required number of events in an interim analysis, N_2 : required number of events in the final analysis

2.1.3. Change in the Required Number of Events in the Analysis for Primary Evaluation

Ramucirumab was approved in Japan in March 2015 and Korea in April 2015. As of March 2017, when the independent data monitoring committee (IDMC) emergency meeting was held, an application was filed for the approval of partial changes to marketing approval matters on the indication of nivolumab for gastric cancer in Japan in December 2016, and early use of nivolumab was expected.

Using these drugs, which were not considered at the time of study initiation, in subsequent therapy may prolong OS and reduce the power assumed before study initiation. The IDMC was therefore consulted for that concern in the IDMC emergency meeting in March 2017 and recommended that the required total number of events be changed to 491 events. Accordingly, the protocol was revised on April 12, 2017, to change the required number of events to 491 in the analysis for the primary evaluation.

2.2. Subject Registration and Randomization Method

2.2.1. Randomization Method

Subjects will be randomized in a 1:1 ratio to TAS-118/L-OHP (study arm, Arm A) or S-1/CDDP (control arm, Arm B).

Treatment assignment will be performed *via* an interactive web response system (IWRS) using a dynamic allocation method (minimization method) stratified by the following allocation adjustment factors:

- Measurable lesion: (yes, no)
- Eastern Cooperative Oncology Group (ECOG) performance status (PS): (0, 1)
- Country: (Japan, Korea)

3. Data to Be Used in Statistical Analysis and Analysis Populations

3.1. Materials and Data Other Than Case Report Forms

A list of materials and data required for statistical analysis other than the case report form (CRF) is presented in Table 3-1.

Table 3-1 Materials and data required for statistical analysis other than the CRF

Parameters	Materials or data	Remarks
Randomization	True allocation key	To identify the treatment randomized to each subject.
Data at randomization	Patient baseline characteristics information, such as the allocation adjustment factors used for patient randomization, that is provided with the true allocation key by the trial registration office	To summarize the information using both data at randomization and CRF data.
Coding to MedDRA terms	Corresponding table between terms of adverse events, medical history, and complications recorded by the investigator and MedDRA terms	To summarize the proportion of subjects with adverse events by using unified terms.
Name of concomitant medications and therapies	Corresponding data between terms of concomitant medications recorded by the investigator and WHO Drug Dictionary	To summarize concomitant medications by using unified terms.
Inclusion of subjects	Flag of the inclusion of each subject in each analysis population	To identify which analysis population each subject is included in.
Significance level in the primary analysis for the primary evaluation	Significance level calculated and information in interim analyses	To calculate the significance level in the primary analysis.
Information on follow-up censoring	Date of follow-up censoring	To reflect the date of follow-up censoring as the date of censoring in survival analysis in subjects whose consent to survival information follow-up cannot be confirmed.
Information on deviations	Nature of deviation and important deviation flag	To provide a list of deviated subjects.
Institutional normal range	Institutional normal range of laboratory values	To determine whether laboratory values are normal or abnormal or evaluate the change in CTCAE grade.

3.2. Analysis Populations

A list of the name and definition of analysis populations is presented in Table 3-2.

If there are any problems with the inclusion of subjects in each analysis population except the criteria listed below, the sponsor will determine in which population the subject should be included before analysis in the discussion with the medical advisor.

Table 3-2 Definitions of analysis populations

Analysis populations	Definition
Enrolled population	All patients enrolled in this trial
As-treated (AT) population	Enrolled patients who received at least one dose of the study drug
Full analysis set (FAS)	Patients in AT population who have advanced gastric cancer at randomization
Tumor response-evaluable population	Patients in FAS population who have measurable lesions (at least one target lesion) at baseline and are evaluated for tumor responses at least once during treatment with the study drug

4. Timing of Analyses and Significance Level

In this study, two interim analyses were planned and implemented before the analysis for the primary evaluation. Based on the results of the two interim analyses and IDMC reviews, the timing of the analysis for the primary evaluation is defined in Table 4-1.

In the analysis for the primary evaluation, whether TAS-118/L-OHP is superior to S-1/CDDP with respect to OS will be tested in the FAS population. As the test method, a stratified log-rank test will be conducted according to the following allocation adjustment factors: measurable lesion (yes, no), ECOG PS (0, 1), and country (Japan, Korea).

Table 4-1 Timing of each analysis and significance level

	Timing of analysis	Significance level at sample size determination
First interim analysis	When the number of death events reaches 100 (Date: April 22, 2016)	Not tested
Second interim analysis	When the number of death events reaches 219 (Date: November 30, 2016)	0.305%
Analysis for the primary evaluation	When the required number of death events, 491, is confirmed	4.899%

These significance levels were calculated using the Lan and DeMets method (O'Brien-Fleming type) at the time of sample size determination. The significance levels to be spent will be recalculated in the interim analyses and the analysis for the primary evaluation. The procedures for recalculation are presented in Section "5.2.4.12 Log-Rank Test Using Recalculated α "

5. Endpoints and Common Considerations for Statistical Analysis

5.1. Endpoints

The definition of efficacy, safety, and compliance is presented below.

Based on the data cutoff date decision record created according to the data cutoff procedures, the range of entry data will be identified. In the primary analysis, the cutoff date is handled as the event date in the 491st subject.

5.1.1. Efficacy Endpoints

The efficacy endpoints in this study are the primary endpoint of OS and the secondary endpoints of PFS, TTF, ORR, and DCR, as well as the duration of overall response, the duration of complete response, and tumor shrinkage rate.

5.1.1.1 OS

Overall survival is defined as the time (in months) from the date of randomization to the date of death due to any cause. In the primary analysis, subjects who are not known to be dead or are alive as of the cutoff date will be censored at the date of the last survival confirmation or the cutoff date, whichever comes first.

The OS cutoff date used for the analysis is defined as the event date in the 491st subject in the FAS population. In the analyses in the enrolled population, the cutoff date is defined as the event date in the 491st subject in the enrolled population.

A list of the definition of OS events is presented in Table 5-1.

Table 5-1 Definition of OS events

	Case	Date
Event	Subjects died before the cutoff date.	Date of death
Censoring	Subjects failed to be followed up for death or are alive as of the cutoff date.	Date of the last survival confirmation* ¹ or the cutoff date, whichever comes first

*1: The latest date among the dates recorded in all pages of the CRF

5.1.1.2 PFS

Progression-free survival is defined as the time (in months) from the date of randomization until the date of the investigator-assessed disease progression or death due to any cause, whichever comes first. Subjects who start to receive subsequent therapy before disease progression assessment will be censored at the date of the last progression-free assessment before subsequent therapy is initiated. A list of the definition of PFS events is presented in Table 5-2.

Subjects who are progression-free and alive as of the cutoff date will be censored at the date of the last progression-free assessment.

Table 5-2 Definition of PFS events

	Case	Date (if multiple cases apply, the earliest date is used)
Event	Progressive disease (PD) before the data cutoff date	Date of PD confirmation
	Death (PD not observed)	Date of death ^{*1}
Censoring	Non-PD or non-death at the time of data cutoff	Date of the last tumor assessment
	Unknown date of PD	Date of the last tumor assessment
	No tumor assessment after randomization	Date of randomization
	Another subsequent therapy for any reason other than PD	Date of the last tumor assessment before the start of subsequent therapy
	Evaluated as not evaluable (NE)	Date of the last tumor assessment immediately before evaluation
	Other	Date of the last tumor assessment

*1: If no tumor is assessed on the date of imaging assessment specified after the date of the last study treatment and a death event occurs before the acceptable limit of the imaging assessment day (the specified assessment day + 1 week), the event will be handled as a PFS event. If a death event occurs after the acceptable limit, the event is not handled as a PFS event.

5.1.1.3 TTF

Time to treatment failure is defined as the time (in months) from the date of randomization until the date of the last study treatment. Subjects who are still on study treatment as of the cutoff date will be censored at the date of the last study treatment or the cutoff date, whichever comes first. A list of the definition of TTF events is presented in Table 5-3.

Table 5-3 Definition of TTF events

	Case	Date
Event	Discontinued study treatment	Date of the last study treatment
Censoring	Other cases	Date of data cutoff

5.1.1.4 Best Overall Response Assessment

Best overall response as per RECIST version 1.1 is the best response recorded from the start of the study treatment until the end of tumor follow-up period. The confirmation of tumor response for CR or PR is not required, but if SD (or non-CR/non-PD) is confirmed as the best overall response, the criteria for SD (or non-CR/non-PD) must be met in a test conducted at least 6 weeks after entry (acceptable within 1 week before and after the scheduled date).

5.1.1.4.1 ORR

Overall response rate is defined as the proportion of subjects with the best overall response of CR or PR in the tumor response-evaluable population.

$$\text{ORR (\%)}: (\text{CR} + \text{PR}) / (\text{Number of subjects in the tumor response-evaluable population}) \times 100$$

5.1.1.4.2 DCR

The assessment of DCR will be parallel that of ORR, with DCR defined as the proportion of subjects with objective evidence of CR, PR, or SD.

$$\text{DCR (\%)}: (\text{CR} + \text{PR} + \text{SD}) / (\text{Number of subjects in the tumor response-evaluable population}) \times 100$$

5.1.1.4.3 Time to Overall Response

Time to overall response is defined as the time (days) from the date of study drug administration to the first overall response of CR or PR. The time to overall response will not be calculated if CR or PR is not evident.

$$\text{Time to overall response: Date of first overall response of CR or PR} - \text{Date of study drug administration} + 1$$

5.1.1.4.4 Time to Complete Response

Time to complete response is defined as the time (days) from the date of study drug administration to the first overall response of CR. The time to complete response will not be calculated if CR is not evident.

$$\text{Time to complete response: Date of first overall response of CR} - \text{Date of study drug administration} + 1$$

5.1.1.4.5 Tumor Shrinkage Rate

The sum of the longest diameters of target lesions at the time of imaging assessment $l = 0, 1, \dots, L$ ($l = 0$ at predose, $l \geq 1$ after the first dose and on and before the start date of subsequent therapy) is defined as R_l . Tumor shrinkage rate at the time of imaging assessment l in each subject is defined by the following expression:

$$\frac{100(R_l - R_0)}{R_0}$$

Maximum tumor shrinkage rate in each subject is defined by the following expression:

$$\text{Min}_{1 \leq l \leq L} \left(\frac{100(R_l - R_0)}{R_0} \right)$$

5.1.2. Safety Endpoints

5.1.2.1 Adverse Events and Adverse Drug Reactions

An adverse event (AE) is any untoward medical condition sign, symptom, or disease that occurs in a subject while participating in this study and does not necessarily have a causal relationship with the study drug.

Treatment-emergent adverse events are AEs that occur from the initiation of any study medication administration and do not necessarily have a causal relationship with the study drug.

Any untoward medical sign, symptom, or disease that occurs outside the period of safety follow-up (30 days after the last dose of study medication or until the start of new antitumor therapy, whichever comes first) is not considered an AE. If any AEs are observed after the safety follow-up period has ended, only those AEs suspected to have a causal relationship with the study drug will be recorded in the CRF.

The analysis of AEs will be based on treatment-emergent adverse events.

Adverse drug reactions in this study are defined as adverse events (including laboratory abnormalities) determined to have a reasonable possibility of a causal relationship with the study drug.

5.1.3. Compliance Endpoints

5.1.3.1 Total Dose

The dose in each cycle $l = 1, 2, \dots, L$ (the cycle of the last study treatment $l = L$) is defined as d_l . The definition of d_l in each arm is presented below.

A (TAS-118)

The d_l in Arm A (TAS-118) is defined by the following formula, where the number of doses of 30, 40, 50, and 60 mg is n_1, n_2, n_3 and n_4 , respectively:

$$d_l = 30n_1 + 40n_2 + 50n_3 + 60n_4$$

If an unspecified dose is administered, the d_l in Arm A (TAS-118) is defined by the following formula, where the dose is A and the number of doses is m :

$$d_l = 30n_1 + 40n_2 + 50n_3 + 60n_4 + \sum A_i m_i$$

A (L-OHP)

The d_l in Arm A (L-OHP) is defined as the actual dose [mg/body] administered in each cycle.

B (S-1)

The d_l in Arm B (S-1) is defined by the following formula, where the number of doses of 40, 50, and 60 mg is n_1, n_2 , and n_3 , respectively:

$$d_l = 40n_1 + 50n_2 + 60n_3$$

If an unspecified dose is administered, the d_l in Arm B (S-1) is defined by the following formula, where the dose is A and the number of doses is m :

$$d_l = 40n_1 + 50n_2 + 60n_3 + \sum A_i m_i$$

B (CDDP)

The d_l in Arm B (CDDP) is defined as the actual dose [mg/body] administered in each cycle.

The total dose is the sum of doses administered from the start of Cycle 1 study treatment through the discontinuation of study treatment and is defined by the following formula:

$$D = \sum_{l=1}^L d_l$$

5.1.3.2 Dose Intensity

The definition of actual dose intensity and relative dose intensity is presented in Tables 5-4 and 5-5, respectively.

Table 5-4 Definition of actual dose intensity

Arm		Definition	Unit
A	TAS-118	(Total dose [mg] / BSA [m ²] at entry) / ((Start date of the last cycle treatment – Start date of Cycle 1 treatment + 14) / 7)	[mg/m ² /week]
	L-OHP	(Total dose [mg] / BSA [m ²] at entry) / ((Start date of the last cycle treatment – Start date of Cycle 1 treatment + 14) / 7)	
B	S-1	(Total dose [mg] / BSA [m ²] at entry) / ((Start date of the last cycle treatment – Start date of Cycle 1 treatment + 35) / 7)	
	CDDP	(Total dose [mg] / BSA [m ²] at entry) / ((Start date of the last cycle treatment – Start date of Cycle 1 treatment + 35) / 7)	

Table 5-5 Definition of relative dose intensity

Arm		Definition	Unit
A	TAS-118	$100 \times \text{Actual dose intensity [mg/m}^2\text{/week]} / ((\text{Initial dose [mg]} \times 2 \times 7 / \text{BSA [m}^2\text{] at entry}) / 2 \text{ [week]})$	[%]
	L-OHP	$100 \times \text{Actual dose intensity [mg/m}^2\text{/week]} / ((\text{Initial dose [mg/body]} / \text{BSA [m}^2\text{] at entry}) / 2 \text{ [week]})$	
B	S-1	$100 \times \text{Actual dose intensity} / ((\text{Initial dose [mg]} \times 2 \times 21 / \text{BSA [m}^2\text{] at entry}) / 5 \text{ [week]})$	
	CDDP	$100 \times \text{Actual dose intensity [mg/m}^2\text{/week]} / ((\text{Initial dose [mg/body]} / \text{BSA [m}^2\text{] at entry}) / 5 \text{ [week]})$	

5.2. Common Considerations for Statistical Analysis

5.2.1. Data Handling

5.2.1.1 Handling of Longitudinal Data

5.2.1.1.1 Calculation of Number of Days and Period

In this statistical analysis plan, the number of days and period will be consistently calculated using the following formulas unless otherwise specified:

- After the reference date Day n = target date – reference date + 1
- Before the reference date Day n = target date – reference date

Accordingly, the start date of treatment is Day 1, the day following the start of treatment is Day 2, and the day before the start of treatment is Day –1. The period is 1 day when the target date is just the reference date and 2 days when the target date is the date following the reference date.

If the month is missing in a date used for the calculation, the number of days and period is regarded as missing. If only the day is missing, it is replaced with the first day before calculation (e.g., April xx, 2016 → April 1, 2016). Only when the date of death is missing (year and month observed) and the date of the last survival confirmation in the same year and month is known, the date of death is regarded as the date following the last survival confirmation.

5.2.1.1.2 Calculation of Number of Years, Months, and Weeks

In this statistical analysis plan, the number of years, months, and weeks will be consistently calculated using the following formulas:

- Number of years = Number of days / 365.25
- Number of months = Number of days / 30.4375
- Number of weeks = Number of days / 7

The number of years, months, or weeks will not be rounded unless indicated.

5.2.1.1.3 Acceptable Time Window

For each evaluation time point, data obtained within the time window specified below will be used. If multiple observations or measurements are available within a time window, data outside the time window will be excluded, and then the data obtained on the day closest to the scheduled day will be used for each parameter of observation or measurement. If measurement days before and after the scheduled day are shifted by the same number of days, the data obtained after the scheduled day will be used.

[Imaging assessment (overall response)] The reference day is the randomization day.

Time point	Scheduled day	Time window
6 weeks after randomization	Day 43	Randomization day \leq measurement day \leq Day 50
$n \times 6$ weeks after randomization	Day ($n \times 42 + 1$)	Day ($n \times 42 - 6$) \leq measurement day \leq Day ($n \times 42 + 8$)

[Sum of diameters of target lesions/laboratory values/tumor markers/vital signs (including height/weight/BSA)]

Time point	Scheduled day	Time window
Baseline	Start date of study treatment	Measurement day \leq start date of study treatment

[Cycles used for the analyses]

Relationship between the start date of treatment in each cycle and the assessment day	Cycle used for the analysis
Start date of treatment in Cycle $x <$ assessment day $<$ start date of treatment in Cycle $x + 1$	Cycle x
Start date of treatment in Cycle $x =$ assessment day and assessed after treatment	Cycle x
Start date of treatment in Cycle $x =$ assessment day and examined before treatment	Cycle $x - 1$ If $x = 1$, then baseline is used.
Start date of treatment in Cycle $x <$ assessment day (with no next cycle) within the cycle decision period*	Cycle x
Start date of treatment in Cycle $x <$ assessment day (with no next cycle) outside the cycle decision period*	After safety follow-up

*Data obtained on the date of the last study treatment + 30 days or the start date of subsequent therapy, whichever comes first (including the day), will be included in the cycle decision period. Data obtained within this period after the date of the last study treatment will be included in the last cycle.

When data are collected in each cycle that is not specified in the list, all cycles except those after safety follow-up will be included in the analysis.

5.2.1.2 Handling of Missing Data and Abnormal Values

No imputation of missing data will be performed. All data accepted at each time point will be used for the analysis. The proportion of subjects with adverse events and other proportions will be calculated using the analysis population as the denominator.

All data measured will be used for the analysis unless the cause of abnormalities is obvious, such as laboratory abnormalities caused by hemolysis during blood collection. If there are abnormalities to be

excluded from the analysis, these values will be identified, and the reason for exclusion will be justified.

5.2.1.3 Handling of Treatment Arms

Analyses in the FAS and the tumor response-evaluable population will be based on the randomized treatment arm in principle.

Analyses in the AT population will be based on the actual treatment given to each subject.

5.2.2. Patient Baseline Characteristics

A list of patient baseline characteristics to be measured at entry is presented in Table 5-6.

For ECOG PS, both data obtained at randomization and data entered in “Vital Signs and ECOG (At the time of Randomization)” of the CRF will be used.

For the presence or absence of measurable lesions, both data obtained at randomization and data entered in “Tumor Imaging (Baseline)” of the CRF will be used. In the data from “Tumor Imaging (Baseline),” a measurable lesion is determined as being present when a target lesion is present, and a measurable lesion is determined as being absent when a target lesion is absent.

The number of metastatic organs is defined as the number of “sites of lesions” recorded in “Target Lesion” and “Non-target Lesion” at predose. However, “Stomach” will not be treated as a metastatic organ because all stomach lesions are primary lesions. The number of metastatic organs will be counted as one if multiple metastases are recorded in the same organ, and regional lymph nodes and distant lymph nodes are counted as the same organ. Metastatic organs recorded as “Other” will also be counted respectively in “Number of metastatic organs.”

If multiple measurements are available before treatment initiation, the most recent data before treatment initiation will be used.

Among the patient characteristic factors listed in Table 5-6, the factors used for efficacy analysis are presented in Table 5-7, and the details of categories of each factor are presented in Table 5-8.

If data are missing, the category will be indicated as “Missing.” Missing data will be excluded from the denominator used to calculate proportions and from statistical tests.

Table 5-6 Characteristic factors

Parameter	Factor
Patient baseline characteristics	Sex, age, race, country, height, weight, BSA, ECOG PS
Disease characteristics	Presence/absence of primary lesion, primary site, histological type, detail of diffuse type, detail of intestinal type, HER2, clinical stage, presence/absence of measurable lesion, sum of diameters of target lesions, category of the number of metastatic organs, metastatic organ category, presence/absence of liver metastasis, peritoneum metastasis, ascites, and lymph node metastasis, JCOG Index*
Prior treatment	Neoadjuvant chemotherapy (yes/no, regimen, detail of chemotherapy), adjuvant chemotherapy (yes/no, regimen, detail of chemotherapy), prior gastrectomy (yes/no, detail of surgery, outcome), other therapy (yes/no)
Laboratory values	Hemoglobin, white blood cell count, neutrophil count, lymphocyte count, platelet count, AST, ALT, ALP, albumin, bilirubin, creatinine, creatinine clearance, Na, K, corrected Ca, CRP, LDH, urine protein, CEA, CA19-9

*JCOG Index: Number (0 to 4) of conditions that apply to a patient among the following risk factors: ECOG PS \geq 1, number of metastatic organs \geq 2, no prior gastrectomy, and ALP deviation from the institutional upper limit of normal (ULN).¹¹⁾

Table 5-7 Factors used for analysis of each characteristic factor

Parameter	Factor
Patient baseline characteristics	Sex, age category 1, age category 2, country, BSA category, ECOG PS
Disease characteristics	Presence/absence of primary lesion, primary site, histological type, detail of diffuse type, detail of intestinal type, HER2, clinical stage, presence/absence of measurable lesion, category of the number of metastatic organs, metastatic organ category, presence/absence of liver metastasis, peritoneum metastasis, ascites, and lymph node metastasis
Prior treatment	Neoadjuvant chemotherapy (yes/no), adjuvant chemotherapy (yes/no), prior gastrectomy (yes/no), other therapy (yes/no)
Laboratory values	CEA category, CA19-9 category

Table 5-8 Details of categories

Category	Details
Sex	Male, female
Age category 1	<65 years, ≥65 years
Age category 2	<75 years, ≥75 years
Country	Japan, Korea
BSA category	<1.250, ≥1.250 to <1.500, ≥1.500
ECOG PS	0, 1
Primary site	Stomach, gastroesophageal junction
Histological type	Diffuse type, intestinal type, unknown
Detail of diffuse type	Poorly differentiated adenocarcinoma, signet-ring cell carcinoma, mucinous adenocarcinoma
Detail of intestinal type	Well-differentiated tubular adenocarcinoma, moderately differentiated tubular adenocarcinoma, papillary adenocarcinoma
HER2	Negative, positive (IHC 3+), positive (IHC 2+ / FISH+), unknown
Clinical stage	Stage IV, recurrent
Category of the number of metastatic organs	1, 2, ≥3
Metastatic organ category	Liver, lung, regional lymph nodes, distant lymph nodes, peritoneum, ascites, pleura, pleural effusion, bone, bone marrow, brain, skin, adrenal glands, esophagus, colon and rectum, ovary, pancreas, other
JCOG Index	Good risk (0 or 1), moderate risk (2 or 3), poor risk (4)
Neoadjuvant/adjuvant chemotherapy	Chemotherapy, chemoradiotherapy, radiotherapy
Detail of chemotherapy	Paclitaxel, Abraxane, docetaxel, irinotecan, S-1, capecitabine, 5-FU, UFT, oxaliplatin, cisplatin, leucovorin, trastuzumab, other
CEA category	≤ institutional ULN, > institutional ULN
CA19-9 category	≤ institutional ULN, > institutional ULN

5.2.3. Statistical Methods

5.2.3.1 Summary Statistics for Continuous Data

Summary statistics for continuous data, if indicated, include the sample size, minimum, maximum, median, first quartile, third quartile, mean, and standard deviation.

5.2.3.2 Proportion

Proportions will be calculated using the following formula:

$$\text{Proportion (\%)} = \frac{\text{Number of relevant subjects}}{\text{Number of subjects in each arm in the analysis population}} \times 100$$

5.2.3.3 Test Methods

(1) Patient baseline characteristics

The test methods used to investigate whether patient baseline characteristics are biased between the treatment arms are presented in Table 5-9.

Table 5-9 Test methods for patient baseline characteristics

Test method	Parameter	Factor
t test (homogeneity of variance)	Patient baseline characteristics	Age, height, weight, BSA
	Disease characteristics	Sum of diameters of target lesions
	Laboratory values	Hemoglobin, white blood cell count, neutrophil count, lymphocyte count, neutrophil/lymphocyte count ratio, platelet count, AST, ALT, ALP, albumin, bilirubin, creatinine, creatinine clearance, Na, K, corrected Ca, CRP, LDH, urine protein, CEA, CA19-9
Fisher's exact test	Patient baseline characteristics	Sex, age category 1, age category 2, country, ECOG PS
	Disease characteristics	Presence/absence of primary lesion, primary site, clinical stage, presence/absence of measurable lesion, presence/absence of liver metastasis, peritoneal metastasis, ascites, and lymph node metastasis
	Prior treatment	Presence/absence of neoadjuvant chemotherapy, presence/absence of adjuvant chemotherapy, prior gastrectomy, other therapy
	Laboratory values	CEA category, CA19-9 category
χ^2 test (Pearson chi-squared test without continuity correction)	Patient baseline characteristics	–
	Disease characteristics	Histological type, HER2
	Prior treatment	–
Wilcoxon test (normal approximation with continuity correction)	Patient baseline characteristics	BSA category
	Disease characteristics	Category of the number of metastatic organs, JCOG Index
	Laboratory values	Urine protein To be tested after it is converted into numbers as follows: –1: – 0: +- 1: + 2: 2+ 3: 3+ 4: 4+

(2) Survival analysis

A stratified log-rank test or log-rank test will be used to compare the distribution of survival times between the treatment arms. The endpoints of OS, PFS, and TTF will be evaluated.

(3) Comparison of proportions

Fisher's exact test will be used to compare the proportion of subjects between the treatment arms. The endpoints of ORR and DCR will be evaluated. For the endpoints of ORR and DCR, a Cochran–Mantel–Haenszel test stratified by the allocation adjustment factors will also be conducted.

5.2.3.4 Model Analysis

(1) Cox proportional hazards model

A Cox proportional hazards model will be used to estimate the hazard ratio. The hazard function at time t is defined as follows:

$$\lambda(t; \mathbf{X}) = \lambda_0(t) \exp(\mathbf{X}\boldsymbol{\beta}),$$

where $\lambda_0(t)$ is a baseline hazard function, \mathbf{X} is a covariate, and $\boldsymbol{\beta}$ is a regression coefficient.

5.2.3.5 Estimation Methods

(1) Confidence interval of proportions

The confidence interval of ORR, DCR, and the proportion of subjects with adverse events will be calculated using the F distribution-based method proposed by Clopper and Pearson (1934).⁸⁾

The upper limit of confidence interval P_U and the lower limit of confidence interval P_L will be estimated using the following formulas, where P is a proportion, X is the number of subjects with events, N is the number of subjects in the analysis population, and α is a two-sided significance level:

$$P = X/N$$

$$P_U = \frac{v_1 F_{v_1, v_2}(\alpha/2)}{v_2 + v_1 F_{v_1, v_2}(\alpha/2)}, v_1 = 2(X + 1), v_2 = 2(N - X)$$

$$P_L = \frac{v_2}{v_2 + v_1 F_{v_1, v_2}(\alpha/2)}, v_1 = 2(N - X + 1), v_2 = 2X$$

(2) Estimation of survival curves

The survival time function $S(t_j)$ will be estimated using the following formula based on the Kaplan–Meier method, where $t_1 < t_2 < \dots < t_N$ is the time of event onset in subject $j = 1, \dots, N$, n_j is the number of subjects at risk at time t_j , and d_j is the number of subjects with events:

$$\hat{S}(t_N) = \prod_{j=1}^N \left(1 - \frac{d_j}{n_j}\right)$$

(3) Confidence interval of MST

The confidence interval of MST will be calculated using the method based on the sign test for censored data.⁹⁾

The standard error of survival time function $S(t_j)$ will be estimated using the following formula:

$$\hat{\sigma}(\hat{S}(t_N)) = \hat{S}(t_N) \sqrt{\sum_{j=1}^N \frac{d_j}{n_j(n_j - d_j)}}$$

The confidence interval of MST is defined as the time of event onset within the following interval, where α is a two-sided significance level:

$$\left\{t: (1 - \hat{S}(t) - 0.5)^2 \leq \chi_1^2(\alpha) \hat{\sigma}^2(\hat{S}(t))\right\}$$

(4) Confidence interval of hazard ratios

The covariance matrix of a regression coefficient estimate $\hat{\boldsymbol{\beta}}$ is expressed using the following formula, where $l(\boldsymbol{\beta})$ is a log-likelihood function:

$$\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}) = - \left[\frac{\partial^2 l(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right]^{-1}$$

Ties in survival time, if any, will be processed using the Breslow method.

The confidence interval of a hazard ratio estimate of the p th covariate, $\exp(\hat{\beta}_p)$, will be estimated using the following expression,¹⁰⁾ where α is a two-sided significance level and $\hat{V}_p(\hat{\boldsymbol{\beta}})$ is the p th diagonal element in the covariance matrix $\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}})$:

$$\exp\left(\hat{\beta}_p \pm Z_{\alpha/2} \sqrt{\hat{V}_p(\hat{\beta})}\right)$$

5.2.4. Sample Code

5.2.4.1 Stratified Log-Rank Test

```
proc lifetest data = TargDS method = KM conftype = linear;
  TIME TimeVAR * CensorVAR(1);
  STRATA StratifiedVARs / Group = TrtVAR;
  ods output HomTests = hom;
run;

/* Explanation of variables (the same applies hereinafter) */
/* TargDS: dataset to be analyzed */
/* TimeVAR: variable with the time to event or time to censoring */
/* CensorVAR: variable that provides event or censoring, in the sample above, 1 = censoring */
/* StratifiedVARs: variables handled as strata or subgroups such as allocation factors, more than one
variable is possible. */
/* TrtVAR: treatment arm [1: A(TAS-118/L-OHP), 0: B(S-1/CDDP)] */
```

5.2.4.2 Log-Rank Test

```
proc lifetest data = TargDS method = KM
  conftype = linear;
  TIME TimeVAR * CensorVAR(1);
  STRATA TrtVAR;
  ods output Quartiles = qua;
  ods output ProductLimitEstimates = est;
run;
```

5.2.4.3 Cox Proportional Hazards Model

```
proc phreg data = TargDS;
  class TrtVAR StratifiedVARs;
  model TimeVAR * CensorVAR(1) = TrtVAR StratifiedVARs
    / r1 TIES = BRESLOW;
  hazardratio TrtVAR / CL = Wald;
run;
```

5.2.4.4 Cox Proportional Hazards Model (by Baseline Characteristic Factor)

```
proc phreg data = TargDS;
  class TrtVAR StratifiedVAR;
  model TimeVAR * CensorVAR(1) = TrtVAR StratifiedVAR TrtVAR * StratifiedVAR
    / r1 TIES = BRESLOW;
  hazardratio TrtVAR / CL = Wald;
run;
```

```
/* More than one StratifiedVAR is prohibited. */
/* To create a forest plot in each Subgroup, use the following results: */
```

```
/* p-value of interaction: p-value of TrtVAR * StratifiedVAR */
/* HR and confidence interval in each Subgroup: HR of each category of StratifiedVAR */
```

5.2.4.5 Restricted Mean Survival Time (RMST)

The boundary time τ is assumed to be 36 months. For the calculation of the number of years, see Section 5.2.1.1.2.

For variance, standard output data from SAS will be used.

For the details of the calculation method, refer to “Measures to evaluate time-to-event responses—To understand restricted mean survival time (RMST),” which is provided by the Japan Pharmaceutical Manufacturers Association.¹¹⁾

```
/* Step1 If survival time is longer than 36 months, then change all subjects to censoring. */
data TargDS2;
  set TargDS;
  if TimeVAR > 36 then CensorVAR = 1;
run;

/* Step2 Calculate RMST */
proc lifetest data = TargDS2 conftype = linear timelim = 36;
  time TimeVAR * CensorVAR(1); /* Day * Censored or not (value at the time of censoring) */
  strata TrtVAR;
  ods output Means = pre_RMST /* Obtain RMST */
            ;
run;

/* Step3 Calculate the confidence interval in each arm and the difference in RMST and its 95%
confidence interval */
data rmst;
  set pre_RMST;
  CI_low = Mean - quantile('NORMAL', .975) * StdErr; /* Lower confidence limit */
  CI_up = Mean + quantile('NORMAL', .975) * StdErr; /* Upper confidence limit */
run;

/* Step 4 Calculate the between-group difference in RMST and its 95% confidence interval */
data rmstdiff;
  merge pre_RMST(where = (TrtVAR = "A(TAS-118/L-OHP)"))
        (keep = TrtVAR Mean StdErr
         rename = (Mean = Mean_A StdErr = StdErr_A))
        pre_RMST(where = (TrtVAR = "B(S-1/CDDP)"))
        (keep = TrtVAR Mean StdErr
         rename = (Mean = Mean_B StdErr = StdErr_B));
  diff = Mean_A - Mean_B; /* Difference in RMST */
  CIdiff_low = diff - quantile('NORMAL', .975) * sqrt(StdErr_A ** 2 + StdErr_B ** 2); /* Lower
confidence limit */
  CIdiff_up = diff + quantile('NORMAL', .975) * sqrt(StdErr_A ** 2 + StdErr_B ** 2); /* Upper
confidence limit */
run;
```

5.2.4.6 95% Confidence Interval of Between-Group Difference in MST by Bootstrap

Method

```
/* Step1 Run 2000 repetitions of sampling with replacement for the number of subjects to be analyzed in each arm. */
```

```
proc surveysselect data = TargDS method = urs rep = 2000
    rate = 1 seed = 20181122 out = out OUTHITS;
strata TRT01P;
run;
proc sort data = out;
    by Replicate;
run;
```

```
/* Step2 Calculate MST in each arm in each repetition */
```

```
proc lifetest data = out method = KM
    conftype = linear;
    TIME AVAL * CNSR(1);
    STRATA TRT01P;
    by Replicate;
    ods output Quartiles = qua;
run;
```

```
/* Step3 Calculate the between-group difference in MST in each repetition */
```

```
data MST;
    merge qua(where = (TRT01P = "A(TAS-118/L-OHP)" and Percent = 50) keep = Replicate
    TRT01P Percent Estimate rename = (Estimate = Estimate_A))
    qua(where = (TRT01P = "B(S-1/CDDP)" and Percent = 50) keep = Replicate TRT01P
    Percent Estimate rename = (Estimate = Estimate_B));
    by Replicate;
    mstdiff = Estimate_A - Estimate_B;
run;
```

```
proc sort data = MST;
    by mstdiff;
run;
```

```
/* Step4 Calculate the 2.5th percentile (lower confidence limit) and 97.5th percentile (upper confidence limit) */
```

```
proc univariate data = MST;
    var mstdiff;
    output out = MST_CI pctlpre = P pctlpts = 2.5 97.5;
run;
```

5.2.4.7 Stratified Cox Proportional Hazards Model

```
proc phreg data = TargDS;
    class TrtVAR StratifiedVARs;
    model TimeVAR * CensorVAR(1) = TrtVAR
        / r1 TIES = BRESLOW;
    strata StratifiedVARs;
```

```
hazardratio TrtVAR / CL = Wald;  
run;
```

5.2.4.8 Comparison of Proportions Between Treatment Arms

```
proc freq data = TargDS;  
  tables StratifiedVARs * TrtVAR / chisq fisher;  
run;
```

5.2.4.9 Confidence Interval of Proportion

```
proc freq data = TargDS;  
  tables KubunVAR / binomial;  
run;  
/* KubunVAR: variable that provides a category in which a proportion is calculated */
```

5.2.4.10 Wilcoxon Test (Two-Sample)

```
proc npar1way data = TargDS Wilcoxon;  
  class TrtVAR;  
  var StratifiedVAR;  
run;
```

Note: To avoid errors in ranking internally processed in SAS, the variance used in VAR statement will be rounded off to the ninth decimal place.

5.2.4.11 Confidence Interval of Median

```
proc univariate data = TargDS cipctldf;  
  var ady;  
run;  
/* Calculate a distribution-free confidence interval using the cipctldf option. */
```

5.2.4.12 Log-Rank Test Using Recalculated α

The log-rank test will be conducted according to the following procedures in this study:

- Using Proc Seqdesign, derive a boundary p-value for the primary evaluation at the time of sample size determination.
- Derive estimated maximum information to establish a boundary at the time of analysis.
- Based on the maximum information as derived above, use Proc Seqdesign to calculate a boundary at the time of each testing.
- In the second interim analysis, use Proc Lifetest to calculate rank-order statistics/covariance matrix in a log-rank test for actual accumulated data.
- Edit the results of Proc Lifetest for Proc Seqtest.
- Using Proc Seqtest, conduct a test based on the boundary calculated by Proc Seqdesign and the edited log-rank test statistics and calculate a boundary used in the analysis for the primary evaluation.
- In the analysis for the primary evaluation, use Proc Lifetest to calculate rank-order statistics/covariance matrix in a log-rank test for actual accumulated data.
- Edit the results of Proc Lifetest for Proc Seqtest.
- Using Proc Seqtest, conduct a test based on the boundary calculated by Proc Seqtest in the second interim analysis and the edited log-rank test statistics.

```
/* Derive a boundary p-value for the primary evaluation at the time of sample size determination. */
```

```

proc seqdesign altref=-0.287682072 errspend boundaryscale = pvalue;
/* The value of altref is the logarithmic value of 0.75, a hazard ratio of TAS-118/L-OHP to S-
1/CDDP as specified in the protocol. */

TwoSidedErrorSpending: design
method(alpha) = errfuncobf /* O'Brien-Fleming type alpha spending function */

info = cum(0.5, 1)
/* The value of info = cum is information fraction. */
/* The value is 0.5 when half of all events are accumulated, */
/* and the value of 1 when all events are accumulated. */

alpha = 0.05 /* Significance level */
nstages = 2 /* Number of tests */
alt = twosided /* Two-sided test */
stop = reject;
samplesize model = twosamplesurvival
(nullhazard = 0.047803254 acctime = 19 foltime = 19);
/* Into t in an exponential distribution  $S(t) = \exp(-\lambda t)$ , */
/* substitute an estimated MST of 14.5 in the S-1/CDDP arm as specified in the protocol to
calculate  $\lambda$ . */
/* Acctime and foltime are the accrual period and the follow-up period, respectively, as defined in
the protocol. */
ods output Boundary = boundpre;
run;

/* Derive estimated maximum information to establish a boundary at the time of analysis. */
data maxinfo;
set boundpre;
where _Stage_ = 2;
theta = log(0.75); /* Estimated logarithmic hazard ratio */
alpha = Bound_LA;
/* Estimated boundary p-value for the primary evaluation */
/* (Value derived by Proc Seqdesign: 0.0244997106 ) */

inva = quantile("normal",1-alpha);
beta =0.15; * power in the primary analysis;
invb = quantile("normal",1-beta);

maxinfo = ((inva + invb) / theta) ** 2; /* Estimated maximum information */
required = maxinfo * 4; /* For reference: required number of events */

call symput("maxinfo",strip(put(maxinfo,best)));
/* Convert the estimated maximum information used in Proc Seqdesign to a macro variable. */
run;

/* Calculate a boundary at the time of analysis using the converted maximum information. */
proc seqdesign altref=-0.287682072 errspend maxinfo = &maxinfo;
TwoSidedErrorSpending: design

```

```
method(alpha) = errfuncobf info = cum(0.5, 1) alpha = 0.05 nstages = 2
alt = twosided stop = reject;
samplesize model = twosamplesurvival
(nullhazard = 0.047803254 acctime = 19 folttime = 19);
ods output ErrSpend = ErrSpend;
ods output Boundary = bound_surv;
run;

/* Stratified log-rank test (the second interim analysis) */
/* MLIWN PSIWN CNTRYIWN means the allocation adjustment factors. */
/* Presence/absence of measurable lesion, ECOG PS, country */
/* TRT01P is an assigned treatment arm. */
proc lifetest data = ADTTE method = KM conftype = linear;
time AVAL * CNSR(1);
strata MLIWN PSIWN CNTRYIWN / Group = TRT01P;
ods output HomStats = outs;
ods output LogrankHomCov = outv;
run;

/* Edit the results of Proc Lifetest for Proc Seqtest. */
data logrank_result;
merge outs outv;
by TRT01P;
keep Variable _Scale_ _Stage_ StdErr Estimate;
Variable = 'Arm';
_Scale_ = "Score";
_Stage_ = 1; /* Designate which number is this test among the planned tests. */
/* Obtain a standard error from the element in the first row and the first column (= variance) of a
2 x 2 covariance matrix for the log-rank test. */
if _n_ = 1 then do;
StdErr = sqrt(A_TAS_118_L_OHP_);
Estimate = LogRank;
output;
end;
run;

/* Conduct a test using */
/* the boundary calculated by Proc Seqdesign and the edited log-rank test statistics. */
proc seqtest boundary = bound_surv boundaryscale = pvalue
parms(testvar = Arm) = logrank_result
errspendadj = errfuncobf /* O'Brien-Fleming type alpha spending function */
order = stagewise;
ods output test = test_surv1; /* Use as a boundary in the analysis for the primary evaluation. */
run;

/* Stratified log-rank test (analysis for the primary evaluation) */
proc lifetest data = ADTTE method = KM conftype = linear;
TIME AVAL * CNSR(1);
```

```

strata MLIWN PSIWN CNTRYIWN / Group = TRT01P;
ods output HomStats = outs2;
ods output LogrankHomCov = outv2;
run;

/* Edit the results of Proc Lifetest for Proc Seqtest. */
data logrank_result2;
  merge outs2 outv2;
  by TRT01P;
  keep Variable _Scale_ _Stage_ StdErr Estimate;
  Variable = 'Arm';
  _Scale_ = "Score";
  _Stage_ = 2; /* Designate which number is this test among the planned tests. */
  /* Obtain a standard error from the element in the first row and the first column (= variance) of a
2 × 2 covariance matrix for the log-rank test. */
  if _n_ = 1 then do;
    StdErr = sqrt(A_TAS_118_L_OHP_);
    Estimate = LogRank;
    output;
  end;
run;

/* Conduct a test using the boundary calculated in the second IDMC */
/* and the log-rank test statistics in the analysis for the primary evaluation. */
ods graphics on;
proc seqtest boundary = test_surv1  boundariescale=pvalue
  parms(testvar=Arm) = logrank_result2
  errspendadj = errfuncobf
  order = stagewise;
ods output test = test_surv2;
run;

```

5.2.4.13 Cochran–Mantel–Haenszel Test

```

proc freq data = TargDS;
  tables StratifiedVARs * TrtVAR * KubunVAR / cmh;
run;
/* KubunVAR: variable that provides a category in which a proportion is calculated */

```

5.2.5. Software to Be Used

The software to be used and the version are presented in Table 5-10.

Table 5-10 Software to be used and version

	Software and version
Operating system	Microsoft Windows 7
Statistical analysis software	SAS 9.4 or later

5.2.6. Dictionaries to Be Used

The dictionaries to be used in analyses are presented in Table 5-11.

Table 5-11 Dictionaries to be used

Term	Dictionary	Remarks
Adverse event term	MedDRA/J	<ul style="list-style-type: none">• System organ class will be coded by MedDRA System Organ Class (SOC) and symptoms by MedDRA Preferred Term (PT).• SOC will be sorted in the order of SOC codes, if presented.• PT will be sorted in the order of PT codes, if presented.• The latest version will be used whenever possible.
Concomitant medication name	WHO Drug Dictionary	<ul style="list-style-type: none">• Medications will be tabulated using Generic Medication Names.• Medications will be sorted in the order of Generic Medication Names.• The latest version will be used whenever possible.

6. Primary Endpoint Analyses

The details of the primary analysis and an overview of its sensitivity analyses and the secondary analyses of the primary endpoint are presented below.

6.1. Primary Analysis for Primary Objective

- The superiority of TAS-118/L-OHP to S-1/CDDP with respect to OS will be tested in the FAS population. As the test method, a stratified log-rank test will be conducted according to the following allocation adjustment factors: ECOG PS (0, 1), measurable lesion (1: yes, 0: no), and country (1: Japan, 0: Korea). If the data on ECOG PS and measurable lesion at randomization differ from the data entered in the CRF, the data obtained at randomization will be used.

6.2. Sensitivity Analyses of Primary Analysis

- In the FAS population, a stratified log-rank test will be conducted according to the allocation adjustment factors entered in the CRF.
- Using a log-rank test without any allocation adjustment factors, the analysis will be conducted.
- In the enrolled population, the primary analysis will be conducted.

6.3. Secondary Analyses of Primary Endpoint

- With the date of observed death as the event date and the date of the last survival confirmation as the censoring date, without consideration of the cutoff date, the same analysis as the primary analysis will be conducted.

The following secondary analyses of the primary endpoint will be conducted in the FAS population:

- The survival curve, MST, and 95% confidence interval will be estimated in each treatment arm. The between-group difference in MST and its 95% confidence interval (by the bootstrap method) will also be calculated. The RMSTs and their 95% confidence intervals will be estimated, with an RMST boundary of 3 years (36 months). In addition, the between-group difference in RMST and its 95% confidence interval will be calculated.
- The number of subjects at risk at 3-month intervals, cumulative event-free rate, and median follow-up period (time to censoring) and its 95% confidence interval will also be presented.
- Using a Cox proportional hazards model with the allocation adjustment factors of study treatment (1: TAS-118/L-OHP, 0: S-1/CDDP), ECOG PS (0, 1) at randomization, measurable lesion (1: yes, 0: no) at randomization, and country (1: Japan, 0: Korea) as covariates and a stratified Cox proportional hazards model according to the allocation adjustment factors, the hazard ratio for each covariate and its confidence interval will be estimated. To estimate the confidence interval of the hazard ratio corresponding to the primary analysis, the confidence level of the stratified Cox proportional hazards model is defined as follows: 100–significance level (%) used for the primary analysis. The 95% confidence interval will also be estimated.
- Using a Cox proportional hazards model with the allocation adjustment factors of study treatment (1: TAS-118/L-OHP, 0: S-1/CDDP), ECOG PS (0, 1) entered in the CRF, measurable lesion (1: yes, 0: no) entered in the CRF, and country (1: Japan, 0: Korea) as covariates and a stratified Cox proportional hazards model according to the allocation adjustment factors, the hazard ratio for each covariate and its 95% confidence interval will be estimated.
- Using a Cox proportional hazards model with study treatment (1: TAS-118/L-OHP, 0: S-1/CDDP) as a covariate, the hazard ratio for the covariate and its 95% confidence interval will also be estimated.
- For each allocation adjustment factor, the survival curve, MST, and 95% confidence interval will be estimated in each treatment arm. The between-group difference in MST and its 95% confidence interval (by the bootstrap method) will also be calculated. The RMSTs and their 95% confidence intervals will be estimated, with an RMST boundary of 3 years (36 months). In addition, the between-group difference in RMST and its 95% confidence interval will be calculated.
- The number of subjects at risk at 3-month intervals, cumulative event-free rate, and median follow-up period (time to censoring) and their 95% confidence intervals will also be presented. Among the allocation adjustment factors, ECOG PS and presence/absence of measurable lesion will be analyzed using both data obtained at randomization and data entered in the CRF.

- For each allocation adjustment factor, using a Cox proportional hazards model with study treatment alone as a covariate and a Cox proportional hazards model with the allocation adjustment factors that have not been used for subgrouping and study treatment as covariates, the hazard ratio between study treatments and its 95% confidence interval will be estimated. Furthermore, the distribution of survival time will be compared between the treatment arms by a log-rank test and a stratified log-rank test according to the allocation adjustment factors that have not been used for subgrouping. Among the allocation adjustment factors, ECOG PS and presence/absence of measurable lesion will be analyzed using both data obtained at randomization and data entered in the CRF.
- For each baseline characteristic factor listed in Table 5-7 in Section “5.2.2 Patient Baseline Characteristics” and each allocation adjustment factor, the MST and its 95% confidence interval will be presented in each treatment arm, and the hazard ratio and its 95% confidence interval will be estimated by a Cox proportional hazards model. The covariate of the model will be study treatment only. The confidence interval of the hazard ratio between the treatment arms will be illustrated in a forest plot, and the p-value of interaction between treatment response and each baseline characteristic factor will be presented to check the uniformity among the strata. The following three covariates will be used to calculate the p-value of interaction in the Cox proportional hazards model: study treatment, baseline characteristic factor, and interaction between study treatment and baseline characteristic factors.
- A breakdown of events and censoring will be tabulated in each treatment arm, and the same analysis will also be conducted by country.
- If the assigned treatment differs from the actual treatment given to a subject, a stratified log-rank test will also be conducted by handling the actual treatment given to the subject as the treatment arm.

7. Secondary Endpoint Analysis

An overview of the main analyses of the secondary endpoints is presented below.

To compare with data from literatures, a confidence coefficient of 95% will be used to present the confidence interval.

7.1. Efficacy Analyses

The following analyses will be conducted for the endpoints of antitumor activity in the tumor response-evaluable population and for PFS and TTF in the FAS population. For the selected endpoints, the same analyses will also be conducted by country.

7.1.1. Antitumor Activity

7.1.1.1 Best Overall Response

- The proportion of subjects with each best overall response will be calculated in each treatment arm.

7.1.1.2 ORR

- The ORR and its 95% confidence interval will be estimated in each treatment arm.
- The ORR will be compared between the treatment arms by Fisher's exact test and Cochran–Mantel–Haenszel test stratified by the allocation adjustment factors. For the allocation adjustment factors, data entered in the CRF will be used.
- For each baseline characteristic factor listed in Table 5-7 in Section “5.2.2 Patient Baseline Characteristics,” the ORR and its 95% confidence interval will be estimated in each treatment arm.
- The ORR analyses specified above will also be conducted by country. In the stratified Cochran–Mantel–Haenszel test, the allocation adjustment factors other than the country will be used as strata.

7.1.1.3 DCR

- The DCR will be analyzed in each treatment arm in the same manner as specified in Section “7.1.1.2 ORR.”

7.1.1.4 Time to Response

- Cycles in which overall response has first been determined to be CR or PR after the start of study treatment will be tabulated, and the summary statistics of the number of days to CR or PR will be presented in each treatment arm.

7.1.1.5 Duration of Overall Response

- The following analyses will be conducted in subjects with the best overall response of CR or PR in the tumor response-evaluable population:
- The survival curve of the duration of overall response, MST, and 95% confidence interval will be estimated in each treatment arm. In addition, the number of subjects at risk at 3-month intervals, cumulative event-free rate, and median follow-up period and their 95% confidence intervals will be presented, and the analyses specified above will also be conducted by country. Duration of complete response
- In subjects with the best overall response of CR in the tumor response-evaluable population, the analyses specified in Section “7.1.1.5 Duration of Overall Response” will be conducted.

7.1.2. PFS

- Progression-free survival will be analyzed in the same manner as specified in Section “6. Primary Endpoint Analyses” (excluding the analyses of RMST and between-group difference in MST). In

the analyses using the allocation adjustment factors, only data entered in the CRF will be used with a confidence level of 95% for all confidence intervals.

- In the enrolled population, the same analyses will be conducted.

7.1.3. TTF

- Time to treatment failure will be analyzed in the same manner as specified in Section “6. Primary Endpoint Analyses” (excluding the analyses in the enrolled population and the analyses of RMST and between-group difference in MST). In the analyses using the allocation adjustment factors, only data entered in the CRF will be used with a confidence level of 95% for all confidence intervals.

7.1.4. Tumor Shrinkage Rate

- The distributions of maximum tumor shrinkage rate and tumor shrinkage rate at 6 and 12 weeks after entry will be illustrated in a waterfall plot in each treatment arm.
- The summary statistics will be presented for the maximum tumor shrinkage rate, tumor shrinkage rates at 6 and 12 weeks after entry and at the last imaging assessment, and tumor shrinkage rate at the last imaging assessment in subjects who discontinued the study because of PD in each treatment arm. The maximum tumor shrinkage rate will be compared between the treatment arms by a Wilcoxon test.
- The proportion of subjects with tumor shrinkage of $\geq 20\%$ at 6 weeks after entry will be presented and compared between the treatment arms by Fisher’s exact test.

7.2. Safety Analysis

The following analyses will be conducted in the AT population. These analyses except listings will also be conducted by country. The listings will be provided in the enrolled population.

7.2.1. Adverse Events and Adverse Drug Reactions

- The number and proportion of subjects with adverse events and adverse drug reactions will be calculated by MedDRA SOC and PT and grade in each treatment arm. For grade 2 or higher events, grade 3 or higher events, and events with an incidence of $\geq 10\%$ in either arm (an incidence of $\geq 5\%$ for adverse drug reactions), the number and proportion of subjects will also be calculated by MedDRA SOC and PT and grade.
- According to ECOG PS entered in the CRF and age (< 75 years), the number and proportion of subjects with adverse events and adverse drug reactions will be calculated by MedDRA SOC and PT and grade in each treatment arm.
- The number and proportion of subjects with adverse events leading to discontinuation (subjects with adverse events recorded in “Treatment Discontinuation” of the CRF) will be calculated by MedDRA SOC and PT in each treatment arm.
- The number and proportion of subjects with adverse events leading to interruption of TAS-118 or S-1 within a cycle will be tabulated by MedDRA PT in each treatment arm.
- The number and proportion of subjects with adverse events leading to the interruption of L-OHP or CDDP will be tabulated by MedDRA SOC and PT in each treatment arm.
- The number and proportion of subjects with adverse events leading to a skip of L-OHP or CDDP will be tabulated by MedDRA SOC and PT in each treatment arm.
- The number and proportion of subjects with adverse events leading to dose reduction in TAS-118, S-1, L-OHP, or CDDP will be tabulated by MedDRA SOC and PT in each treatment arm.
- The number and proportion of subjects with adverse events leading to prolongation of the TAS-118 or S-1 recovery period will be tabulated by MedDRA SOC and PT in each treatment arm.
- The number and proportion of subjects with serious adverse events will be tabulated by MedDRA SOC and PT in each treatment arm. Serious adverse drug reactions will also be analyzed in the same manner.
- A list of adverse event type, grade, time of onset, causality, course, treatment, and other information will be created in the enrolled population.
- For serious adverse events, a list of grade, outcome, date of outcome, seriousness, causality, and other information will be created in the enrolled population.

7.2.2. Deaths

- In subjects in the enrolled population who died within 90 days after entry or before the end of the safety follow-up period, a list will be created for the reason for death, start date of treatment, end date of treatment, date of death, number of days from entry to death, number of days from treatment completion to death, and other information.

7.3. Composition of Analysis Populations

The following analyses will be conducted in all subjects. These analyses except listings will also be conducted by country.

- The inclusion of subjects in each analysis population will be tabulated in each treatment arm. The reason for exclusion will also be tabulated.

7.4. Deviation

The following analyses will be conducted in the enrolled population. These analyses except listings will also be conducted by country.

- The presence or absence and nature of deviation will be tabulated in each treatment arm.

7.5. Analyses of Patient Baseline Characteristics

The following analyses will be conducted in the AT and FAS populations. These analyses except listings will also be conducted by country.

- For the baseline characteristic factors listed in Table 5-6 of Section “5.2.2 Patient Baseline Characteristics,” the number and proportion of subjects and summary statistics will be presented in each treatment arm.

7.6. Compliance Analyses

The following analyses will be conducted in the AT and FAS populations. These analyses except listings will also be conducted by country.

7.6.1. Dose Intensity

- The summary statistics of actual dose intensity and relative dose intensity and the number and proportion of subjects with each category of relative dose intensity (<80%, ≥80% to <90%, ≥90%) will be presented in each treatment arm.

7.6.2. Compliance

- The number and proportion of subjects with dose reduction, interruption, skip (only for L-OHP and CDDP), and prolongation of the recovery period (only for TAS-118 and S-1) will be calculated in each treatment arm. These analyses will be conducted for each cycle and for the overall cycles.

7.6.3. Discontinuation

- The presence or absence of study discontinuation and the reason for discontinuation will be tabulated in each treatment arm.
- A list of adverse events leading to discontinuation will be created in the enrolled population.

7.6.4. Cumulative Dose/Duration of Treatment

- The summary statistics of the total duration of treatment and cumulative dose will be presented by treatment arm and study drug. The total duration of TAS-118 and S-1 treatment will be calculated by dividing the sum of the number of doses by 2; the total duration of L-OHP and CDDP treatment will be the sum of the number of doses.

7.6.5. Treatment Completion

- The number of subjects who completed the treatment in each cycle will be tabulated by treatment arm. Subjects who received doses of each drug that are not less than the doses specified below in one cycle are regarded as those who completed the treatment and otherwise as those who did not complete the treatment. The doses required for completion are presented in Table 7-1.

Table 7-1 Doses required for completion

TAS-118	14 packages
S-1	84 capsules (at a dose of 40 or 50 mg at the start of cycle) 126 capsules (at a dose of 60 mg at the start of cycle)
L-OHP and CDDP	1 dose

7.6.6. Treatment Cycle

- The summary statistics of the number of subjects in each cycle and the last cycle number in each subject will be presented by treatment arm.

7.6.7. Proportion of Outpatient Visits

- The summary statistics of the proportion of outpatient visits will be presented by treatment arm. The proportion of outpatient visits will be calculated as follows. In this calculation, the cutoff date is defined as the data cutoff date recorded in the data cutoff date decision record. The definition of hospitalization and the last dose is presented in Table. 7-2.

$$\text{Proportion of outpatient visits: } 100\% - \frac{\text{Duration of hospitalization}}{\text{Duration of treatment}} \times 100$$

Duration of hospitalization: End date of hospitalization – Start date of hospitalization + 1
(If a subject experienced more than one hospitalization, add up the durations of hospitalization)

Duration of treatment: Date of the last treatment – Start date of Cycle 1 treatment + 1

Table 7-2 Definition of hospitalization and last treatment

Start date of hospitalization	If a subject is hospitalized before the start date of Cycle 1 treatment, the duration is defined as the time from the start date of Cycle 1 treatment.
End date of hospitalization	In a subject who discontinued the treatment and is still hospitalized after the date of the last treatment, the date of the last treatment is defined as the end date of hospitalization. In a subject who continues to receive the study drug as of the data cutoff date, the data cutoff date is defined as the end date of hospitalization.
Date of the last treatment	In a subject who continues to receive the study drug as of the data cutoff date, the data cutoff date is defined as the date of the last treatment.

7.7. Concomitant Medications

- The use of concomitant medications will be tabulated by Generic Medication Name of WHO Drug Dictionary in each treatment arm.

7.8. Subsequent Therapy

- The number and proportion of subjects who used subsequent therapies listed in Table 7-3 will be tabulated in each treatment arm. This analysis will also be conducted by country.

Table 7-3 Subsequent therapy

Category	Details
Subsequent therapy	No, yes
Type of subsequent therapy	Chemotherapy, chemoradiotherapy, radiotherapy, immune/hormonal therapy, surgery, other
Detail of chemotherapy	Paclitaxel, Abraxane, docetaxel, irinotecan, S-1, capecitabine, 5-FU, UFT, oxaliplatin, cisplatin, leucovorin, trastuzumab, ramucirumab, nivolumab, other
Intent of surgery	Palliative, curative

7.9. Laboratory Values

The following analyses will be conducted in the AT population. These analyses except listings will also be conducted by country. The listings will be based on the enrolled population.

- The change in laboratory values over time from the start of each cycle in each subject will be graphed by treatment arm.
Laboratory values to be evaluated are hemoglobin, white blood cell count, neutrophil count, lymphocyte count, and platelet count.
- The minimum level of laboratory tests at predose and at each cycle in each subject will be identified by treatment arm, the summary statistics will be presented, and the levels will be summarized by grade.
Laboratory tests to be evaluated are hemoglobin, white blood cell count, neutrophil count, lymphocyte count, platelet count, albumin, Na, K, and corrected Ca.
- The maximum level of laboratory tests at predose and at each cycle in each subject will be identified by treatment arm, the summary statistics will be presented, and the levels will be summarized by grade.
Laboratory tests to be evaluated are AST, ALT, ALP, bilirubin, creatinine, Na, K, and corrected Ca.
- The worst level of urine protein at predose and at each cycle in each subject will be identified by treatment arm and summarized by category.
- The minimum level of laboratory tests throughout the study period in each subject will be identified by treatment arm, and the summary statistics will be presented for predose level, minimum level, the number of days to the minimum level, and cycle with the minimum level.
Laboratory tests to be evaluated are hemoglobin, white blood cell count, neutrophil count, and platelet count.
- The summary statistics will be presented for the minimum level of laboratory tests in subjects with grade 3 or higher reduction, the number of days from the start date of onset cycle to the minimum level, and the number of days required for recovery to grade 2 or less from the minimum level within an onset cycle in each treatment arm.
Laboratory tests to be evaluated are hemoglobin, white blood cell count, neutrophil count, and platelet count.
If grade 3 or higher reduction is observed in more than one cycle in the same subject, both decreased levels will be used for calculating the summary statistics.

Subjects who meet any of the following criteria will be regarded as those who are not recovered:

- After grade 3 or higher reduction, grade 2 or less is not measured within the same cycle.
- Measurement date of grade 2 or less reduction in laboratory values – date of the last study treatment ≥ 30 days

When the start date of subsequent therapy \leq date of the last study treatment + 30 days is met, the following criterion will be used:

Measurement date of grade 2 or less reduction in laboratory values \geq start date of subsequent therapy

- After “grade 3 or higher reduction” is replaced with “grade 2 or higher reduction” and “recovery to grade 2 or less” is replaced with “recovery to grade 1 or less,” the same analyses as specified above will be conducted.

7.10. Clinical Findings

The following analyses will be conducted in the AT population. These analyses except listings will also be conducted by country. The listings will be based on the enrolled population.

- The summary statistics will be presented for the number of days from the start of Cycle 1 treatment to the first onset of clinical findings of grade 1 or higher, the number of days to the worst grade, the number of days from the first onset to the outcome of recovery, and the number of days from clinical findings of grade 3 or higher to recovery to grade 2 or less in each treatment arm.
Clinical findings to be evaluated are stomatitis, anorexia, diarrhea, peripheral sensory neuropathy, nausea, and fatigue.
- A list will be created for the time of onset of the clinical findings to be analyzed in each subject, the time of worst grade, the time of grade 3 or higher, and the time required for recovery in each treatment arm.

7.11. Tumor Markers

The following analyses will be conducted in the AT population. These analyses except listings will also be conducted by country. The listings will be based on the enrolled population.

- The change in tumor markers over time will be graphed in each treatment arm.

7.12. Vital Signs and ECOG PS

The following analyses will be conducted in the AT population. These analyses except listings will also be conducted by country. The listings will be based on the enrolled population.

- The summary statistics of measured weight and change from baseline at the time of discontinuation will be presented in each treatment arm.
- A cross table of ECOG PS will be created at baseline and discontinuation in each treatment arm.

8. Review and Revision of Statistical Analysis Plan

This statistical analysis plan will be finalized as version 1 before the receipt of CRF data or a true allocation key used to create the results of statistical analysis. If a change or addition is required after the finalization of version 1, a revised version will be prepared after the background, and the reason for changes are clarified. Newly planned analyses will be separated from the preplanned analyses and included in the change record.

9. Development of Output Plan

This statistical analysis plan provides the representative analyses only. All analyses to be conducted are presented in a separate output plan.

10. References

- 1) 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;45:228-247.

2) [REDACTED]

- 3) [REDACTED]
- 4) Ryu MH, Baba E, Lee KH, Boku N, Park YI, Hyodo I, et al. Phase III trial of a 3-weekly versus 5-weekly schedule of S-1 plus cisplatin (SP) combination chemotherapy for first-line treatment of advanced gastric cancer (AGC): SOS study. *J Clin Oncol*. 2013;31 (suppl; abstr LBA4024).
- 5) Lan KKG, DeMets DL. Discrete Sequential Boundaries for Clinical Trials. *Biometrika*. 1983;70:659-663.
- 6) Schoenfeld D. The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika*. 1981;68:316-319.
- 7) Shuster JJ. Power and Sample Size for Phase III Clinical Trials of Survival. In: Crowley J, Pauler Ankerst D, editors. *Handbook of Statistics in Clinical Oncology*. 2nd ed. Boca Raton: Chapman & Hall/CRC. 2006;289-333.
- 8) Clopper CJ, Pearson ES. The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika*. 1934;26:404-413.
- 9) Brookmeyer R, Crowley J. A Confidence Interval for the Median Survival Time. *Biometrics*. 1982;38:29-41.
- 10) Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. 2nd ed. New York: Wiley. 2002;99-104.
- 11) Takahari D, Boku N, Mizusawa J, Takashima A, Yamada Y, Yoshino T, et al. Determination of prognostic factors in Japanese patients with advanced gastric cancer using the data from a randomized controlled trial. *Oncologist*. 2014;19:358-366.
- 12) Summary measures to evaluate time-to-event responses—To understand restricted mean survival time (RMST)— (in Japanese). Available from: <http://www.jpma.or.jp/medicine/shinyaku/tiken/allotment/pdf/rmst.pdf>.