STATISTICAL ANALYSIS PLAN

STUDY TITLE: **A PHASE III, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED STUDY OF ATEZOLIZUMAB PLUS TIRAGOLUMAB IN COMBINATION WITH PACLITAXEL AND CISPLATIN COMPARED WITH PACLITAXEL AND CISPLATIN AS FIRST-LINE TREATMENT IN PATIENTS WITH UNRESECTABLE LOCALLY ADVANCED, UNRESECTABLE RECURRENT, OR METASTATIC ESOPHAGEAL SQUAMOUS CELL CARCINOMA**

STUDY NUMBER: YO42138

STUDY NAME: SKYSCRAPER-08

VERSION NUMBER: 1

ROCHE COMPOUND(S): Tiragolumab (RO7092284)
Atezolizumab (RO5541267)

EUDRACT NUMBER: Not Applicable

IND NUMBER: Not Applicable

NCT NUMBER: NCT04540211

PLAN PREPARED BY: [redacted], Ph.D

Date and Time(UTC)	Reason for Signing	Name
04-Mar-2022 04:47:07	Company Signatory	[redacted]

STATISTICAL ANALYSIS PLAN APPROVAL

SPONSOR: F. Hoffmann-La Roche Ltd
LEGAL REGISTERED ADDRESS: Grenzacherstrasse 124
4070 Basel, Switzerland

DATE FINAL: See electronic date stamp above

CONFIDENTIAL

This is an F. Hoffmann-La Roche Ltd document that contains confidential information. Nothing herein is to be disclosed without written consent from F. Hoffmann-La Roche Ltd.

STATISTICAL ANALYSIS PLAN VERSION HISTORY

This Statistical Analysis Plan (SAP) was developed based on Roche SAP model document Version 2, 26 October 2020.

SAP Version	Approval Date	Based on Protocol (Version, Approval Date)
1	see electronic date stamp on title page	Version 5, 21 January 2022

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation or Term	Description
ADA	anti-drug antibody
AE	adverse event
AESI	adverse event of special interest
[redacted]	[redacted]
Atezo+Tira+PC	atezolizumab plus tiragolumab in combination with paclitaxel and cisplatin
[redacted]	[redacted]
CI	confidence interval
CR	complete response
[redacted]	[redacted]
DOR	duration of response
ECOG	Eastern Cooperative Oncology Group
EORTC	European Organisation for Research and Treatment of Cancer
ESCC	esophageal squamous cell carcinoma
FAS	full analysis set
GHS/QoL	global health status and quality of life
HR	hazard ratio
IA	interim analysis
iDMC	independent Data Monitoring Committee
IRF	Independent Review Facility
IxRS	interactive voice or Web-based response system
MDD	minimum detectable difference
NCI CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
NPT	non-protocol anti-cancer therapy
ORR	objective response rate
OS	overall survival
PD-L1	programmed death-ligand 1
PFS	progression-free survival
PK	pharmacokinetic
Placebo+PC	atezolizumab placebo plus tiragolumab placebo in combination with paclitaxel and cisplatin
PR	partial response
QLQ-C30	Quality-of-Life Questionnaire Core 30

QLQ-OES18	Quality of Life Questionnaire - Oesophageal Cancer Module
RECIST v1.1	Response Evaluation Criteria in Solid Tumors, Version 1.1
SAP	Statistical Analysis Plan
[redacted]	[redacted]
TTCD	time to confirmed deterioration
[redacted]	[redacted]

1. INTRODUCTION

This Statistical Analysis Plan (SAP) provides details of the planned analyses and statistical methods for Study YO42138 (SKYSCRAPER-08), a Phase III, randomized, double-blind, placebo-controlled study of atezolizumab plus tiragolumab in combination with paclitaxel and cisplatin compared with paclitaxel and cisplatin as first-line treatment in patients with unresectable locally advanced, unresectable recurrent, or metastatic esophageal squamous cell carcinoma (ESCC). The background for the study can be found in the study protocol.

1.1 OBJECTIVES, ENDPOINTS AND ESTIMANDS

Study YO42138 evaluates the efficacy and safety of atezolizumab plus tiragolumab in combination with paclitaxel and cisplatin (hereinafter referred to as Atezo+Tira+PC) compared with atezolizumab placebo plus tiragolumab placebo in combination with paclitaxel and cisplatin (hereinafter referred to as Placebo+PC) as first-line treatment in patients with unresectable locally advanced, unresectable recurrent, or metastatic ESCC. Specific objectives and corresponding endpoints for the study are outlined in Table 1.

The term "study treatment" refers to all protocol-mandated treatments assigned to patients as part of this study and includes atezolizumab/placebo, tiragolumab/placebo, paclitaxel and cisplatin during the induction phase; atezolizumab/placebo and tiragolumab/placebo during the maintenance phase.

Table 1 Objectives and Corresponding Endpoints

Primary Efficacy Objective	Corresponding Endpoints
• To evaluate the efficacy of Atezo+Tira+PC compared with Placebo+PC	• OS, defined as the time from randomization to death from any cause • PFS, defined as the time from randomization to the first occurrence of disease progression or death from any cause (whichever occurs first), as determined by an IRF according to RECIST v1.1
Secondary Efficacy Objective	**Corresponding Endpoints**
• To evaluate the efficacy of Atezo+Tira+PC compared with Placebo+PC	• PFS as determined by the investigator according to RECIST v1.1 • Confirmed ORR, defined as the proportion of patients with a CR or PR on two consecutive occasions ≥4 weeks apart, as determined by an IRF according to RECIST v1.1 • Confirmed ORR as determined by the investigator according to RECIST v1.1 • DOR, defined as the time from the first occurrence of a confirmed objective response to the first occurrence of disease progression or death from any cause (whichever occurs first), as determined by an IRF according to RECIST v1.1 • DOR as determined by the investigator according to RECIST v1.1 • TTCD in patient-reported physical functioning, role functioning, and GHS/QoL as measured by the respective scales of the EORTC QLQ-C30 and defined as the time from randomization to first deterioration (decrease from baseline of ≥10 points) that is either maintained for two consecutive assessments or followed by death from any cause within 3 weeks • TTCD in patient-reported dysphagia as measured by the dysphagia scale of the EORTC QLQ-OES18 and defined as the time from randomization to first deterioration (increase from baseline of ≥10 points) that is either maintained for two consecutive assessments or followed by death from any cause within 3 weeks
[redacted]	

Table 1 Objectives and Corresponding Endpoints (cont.)

	• [redacted]
Safety Objective	**Corresponding Endpoint**
• To evaluate the safety of Atezo+Tira+PC compared with Placebo+PC	• Incidence and severity of adverse events Severity for all events will be graded according to NCI CTCAE v5.0, [redacted]
Pharmacokinetic Objective	**Corresponding Endpoint**
• To characterize the PK profiles of tiragolumab and atezolizumab when given in combination with paclitaxel and cisplatin	• Serum concentration of tiragolumab and atezolizumab at specified timepoints
Immunogenicity Objective	**Corresponding Endpoint**
• To evaluate the immune response to tiragolumab and atezolizumab when given in combination with paclitaxel and cisplatin	• Prevalence of ADAs to tiragolumab at baseline and incidence of ADAs to tiragolumab during the study • Prevalence of ADAs to atezolizumab at baseline and incidence of ADAs to atezolizumab during the study
[redacted]	
Health Status Utility Objectives	**Corresponding Endpoint**
• To evaluate health status utility scores of patients treated with Atezo+Tira+PC compared with Placebo+PC	• Mean change from baseline in the index-based and VAS scores of the EuroQol EQ-5D-5L

Table 1 Objectives and Corresponding Endpoints (cont.)

ADA = anti-drug antibody; [redacted]
Atezo+Tira+PC = atezolizumab plus tiragolumab in combination with paclitaxel and cisplatin;
CR = complete response; [redacted] DOR = duration of response;
EORTC = European Organisation for Research and Treatment of Cancer; GHS/QoL = global health status and quality of life; IRF = Independent Review Facility; NCI CTCAE v5.0 = National Cancer Institute Common Terminology Criteria for Adverse Events, Version 5.0; ORR = objective response rate; OS = overall survival; [redacted]
PFS = progression-free survival; PR = partial response; PK = pharmacokinetic;
Placebo+PC = atezolizumab placebo plus tiragolumab placebo in combination with paclitaxel and cisplatin; QLQ-C30 = Quality-of-Life Questionnaire Core 30; QLQ-OES18 = Quality of Life Questionnaire - Oesophageal Cancer Module; RECIST v1.1 = Response Evaluation Criteria in Solid Tumors, Version 1.1; TTCD = time to confirmed deterioration; [redacted]
VAS = Visual Analog Scale.

1.1.1 Expression of Objectives and Endpoint using the Estimand Framework

Primary endpoints and selected secondary endpoints are expressed using the estimand framework in Table 2, following the International Council on Harmonization E9 (R1) statistical principles for clinical trials (ICH 2020).

Table 2 Objectives and Estimands

Primary Objective(s)	Estimand Definition
• To evaluate the efficacy of Atezo+Tira+PC compared with Placebo+PC	The estimand for **OS** is defined as: • Population: Patients with unresectable locally advanced, unresectable recurrent, or metastatic ESCC • Variable: Time from randomization to death from any cause • Treatment: – Experimental: Atezolizumab 1200 mg followed by Tiragolumab 600 mg, Paclitaxel 175 mg/m^2 and Cisplatin 60–80 mg/m^2 (Induction: Cycles 1–6), followed by Atezolizumab 1200 mg and Tiragolumab 600 mg (Maintenance: Cycles ≥7), administered by IV infusion on Day 1 of each 21-day cycle, until disease progression, loss of clinical benefit, or unacceptable toxicity – Control: Atezolizumab placebo followed by Tiragolumab placebo, Paclitaxel 175 mg/m^2 and Cisplatin 60–80 mg/m^2 (Induction: Cycles 1–6), followed by Atezolizumab placebo and Tiragolumab placebo (Maintenance: Cycles ≥7), administered by IV infusion on Day 1 of each 21-day cycle, until disease progression, loss of clinical benefit, or unacceptable toxicity • Intercurrent events and handling strategies: – Start of NPT prior to the event of interest: treatment policy strategy – Early discontinuation from study treatment for any reason prior to the event of interest: treatment policy strategy • Population-level summary: HR for the variable
	The estimand for **IRF-assessed PFS** is defined similarly as for OS, except: • Variable: Time from randomization to the first occurrence of disease progression or death from any cause (whichever occurs first), as determined by an IRF according to RECIST v1.1

Table 2 Objectives and Estimands (cont.)

Selected Secondary Objective(s)	Estimand Definition
• To evaluate the efficacy of Atezo+Tira+PC compared with Placebo+PC	The estimand for **investigator-assessed PFS** is defined similarly as for IRF-assessed PFS, except: • Variable: Time from randomization to the first documented disease progression as determined by the investigator with the use of RECIST v1.1 or death from any cause, whichever occurs first
	The estimand for **IRF-assessed ORR** is defined similarly as for PFS, except: • Variable: Whether patients achieved a confirmed objective response (i.e., CR or PR on two consecutive occasions ≥4 weeks apart), as determined by an IRF according to RECIST v1.1 • Population-level summary: difference in proportions
	The estimand for **investigator-assessed ORR** is defined similarly as for IRF-assessed ORR, except: • Variable: Whether patients achieved a confirmed objective response (i.e., CR or PR on two consecutive occasions ≥4 weeks apart), as determined by the investigator according to RECIST v1.1
• To evaluate the efficacy of Atezo+Tira+PC compared with Placebo+PC	The estimand for **TTCD** is defined similarly as for PFS, except: • Variable: Time from randomization to first confirmed deterioration on physical functioning, role functioning and GHS/QoL as measured by the respective scales of the EORTC QLQ-C30, and on dysphagia as measured by the respective scale of the EORTC QLQ-OES18; confirmed deterioration is defined as a decrease from baseline of ≥10 points (for the physical functioning, role functioning and GHS/QoL scales) or an increase from baseline of ≥10 points (for the dysphagia scale) that is either maintained for two consecutive assessments or followed by death from any cause within 3 weeks • Intercurrent events and handling strategies: – Start of NPT prior to a confirmed deterioration: treatment policy strategy – Early discontinuation from study treatment for any reason prior to a confirmed deterioration: treatment policy strategy – Death prior to a confirmed deterioration: while alive strategy

Table 2 Objectives and Estimands (cont.)

Atezo+Tira+PC = atezolizumab plus tiragolumab in combination with paclitaxel and cisplatin; CR = complete response; ESCC = esophageal squamous cell carcinoma; EORTC = European Organisation for Research and Treatment of Cancer; GHS/QoL = global health status and quality of life; HR = Hazard Ratio; IRF = Independent Review Facility; IV = intravenous; NPT = non-protocol anti-cancer therapy; ORR = objective response rate; OS = overall survival; PFS = progression-free survival; PR = partial response; Placebo+PC = atezolizumab placebo plus tiragolumab placebo in combination with paclitaxel and cisplatin; QLQ-C30 = Quality-of-Life Questionnaire Core 30; QLQ-OES18 = Quality of Life Questionnaire - Oesophageal Cancer Module; RECIST v1.1 = Response Evaluation Criteria in Solid Tumors, Version 1.1; TTCD = time to confirmed deterioration.

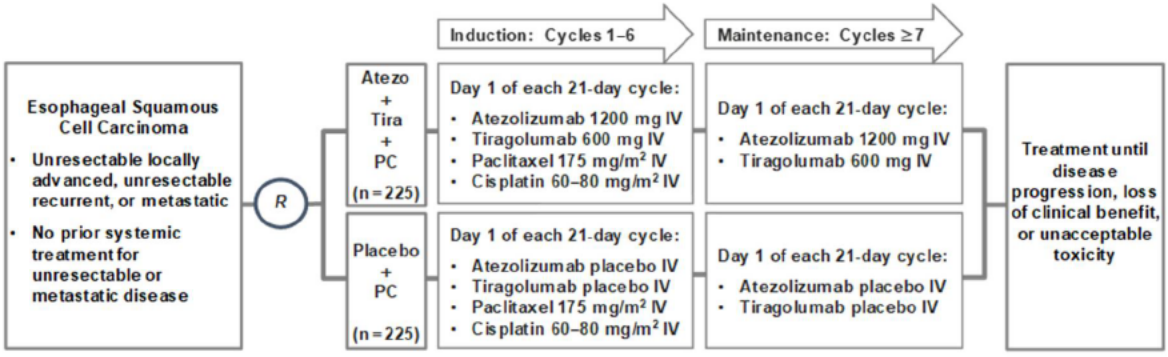
1.2 STUDY DESIGN

This is a Phase III, randomized, multicenter, double-blind study designed to evaluate the efficacy and safety of Atezo+Tira+PC compared with Placebo+PC as first-line treatment in patients with unresectable locally advanced, unresectable recurrent, or metastatic ESCC.

Patients who do not initially meet all eligibility criteria for participation in this study (screen failure) may qualify for 1 re-screening opportunity (for a total of 2 screenings per patient) at the discretion of the investigator. The Medical Monitor is available to advise as needed.

This study will enroll approximately 450 patients, who will be randomized in a 1:1 ratio to the Atezo+Tira+PC arm or the Placebo+PC arm. Eligible patients will be stratified by programmed death-ligand 1 (PD-L1) expression as assessed by a central laboratory through use of the [redacted] previous curative treatment (consisting of either esophagostomy or chemoradiotherapy) (yes vs. no), and Eastern Cooperative Oncology Group (ECOG) Performance Status (0 vs. 1). The study schema is shown in Figure 1.

Figure 1 Study Schema



Atezo+Tira+PC = atezolizumab plus tiragolumab in combination with paclitaxel and cisplatin; IV = intravenous; Placebo+PC = atezolizumab placebo plus tiragolumab placebo in combination with paclitaxel and cisplatin; R = randomization.

Patients will receive study treatment until unacceptable toxicity or loss of clinical benefit as determined by the investigator after an integrated assessment of radiographic and biochemical data, local biopsy results (if available), and clinical status (e.g., symptomatic deterioration such as pain secondary to disease). Because of the possibility of an initial increase in tumor burden caused by immune-cell infiltration in the setting of a T-cell response (termed pseudoprogression) with cancer immunotherapy, radiographic progression per Response Evaluation Criteria in Solid Tumors, Version 1.1 (RECIST v1.1) may not be indicative of true disease progression. In the absence of unacceptable toxicity, patients who meet criteria for disease progression per RECIST v1.1 while receiving study treatment will be permitted to continue treatment if they meet all of the following criteria:

- Evidence of clinical benefit, as determined by the investigator following a review of all available data
- Absence of symptoms and signs (including laboratory values, such as new or worsening hypercalcemia) indicating unequivocal progression of disease
- Absence of decline in ECOG Performance Status that can be attributed to disease progression
- Absence of tumor progression at critical anatomical sites (e.g., leptomeningeal disease) that cannot be managed by protocol-allowed medical interventions

1.3 TREATMENT ASSIGNMENT AND BLINDING

This is a randomized, double-blind study. After written informed consent has been obtained, all screening procedures and assessments have been completed, and eligibility has been established for a patient, the study site will obtain the patient's identification number and treatment assignment from the interactive voice or Web-based response system (IxRS).

Patients will be randomly assigned to one of two treatment arms: Atezo+Tira+PC or Placebo+PC. Randomization will occur in a 1:1 ratio through use of a permuted-block randomization method to ensure a balanced assignment to each treatment arm. Randomization will be stratified according to the following criteria:

- PD-L1 expression as assessed by a central laboratory through use of the [redacted]
- Previous curative treatment (yes vs. no)
- ECOG Performance Status (0 vs. 1)

Study site personnel, patients, and the Independent Review Facility (IRF) will be blinded to treatment assignment during the study. The Sponsor and its agents will also be blinded to treatment assignment, with the exception of individuals who require access to patient treatment assignments to fulfill their job roles during a clinical trial. These roles include the unblinding group responsible, clinical supply chain managers, sample

handling staff, operational assay group personnel, IxRS service provider, and independent Data Monitoring Committee (iDMC) members.

1.3.1 Independent Review Facility

An IRF will be used to enable centralized, independent reviews of tumor assessment images and other clinical data used for assessment of ESCC (e.g., histopathology, tumor markers). IRF reviews will be performed prior to the pre-specified efficacy analyses. IRF membership and procedures will be detailed in an IRF charter.

1.3.2 Data Monitoring

An iDMC will evaluate safety data during the study. Sponsor affiliates will be excluded from iDMC membership. The iDMC will follow a charter that outlines the iDMC roles and responsibilities.

Unblinded safety data will be reviewed by the iDMC after a minimum of 16 patients have been enrolled and have received 2 cycles of study treatment. Subsequently, the iDMC will review safety data approximately every 6 months until the study is unblinded or the study is terminated by the Sponsor. All summaries and analyses for the iDMC review will be prepared by an independent Data Coordinating Center (iDCC).

After reviewing the data, the iDMC will provide a recommendation to the Sponsor as described in the iDMC charter. Final decisions will rest with the Sponsor.

Any outcomes of these data reviews that affect study conduct will be communicated in a timely manner to the investigators for notification of their respective Institutional Review Boards (IRBs)/ Ethics Committees (ECs).

2. STATISTICAL HYPOTHESES

The purpose of this study is to test the hypothesis that Atezo+Tira+PC prolongs the duration of overall survival (OS) and/or progression-free survival (PFS) relative to Placebo+PC in the full analysis set (FAS), defined as all randomized patients regardless of whether they receive any study treatment (see Section 4).

The null (H_0) and alternative (H_1) hypotheses regarding OS or PFS in the FAS can be phrased in terms of the population hazard ratio (HR) λ between the experimental arm and the control arm:

$$H_0: \lambda = 1 \text{ versus } H_1: \lambda \neq 1$$

3. SAMPLE SIZE DETERMINATION

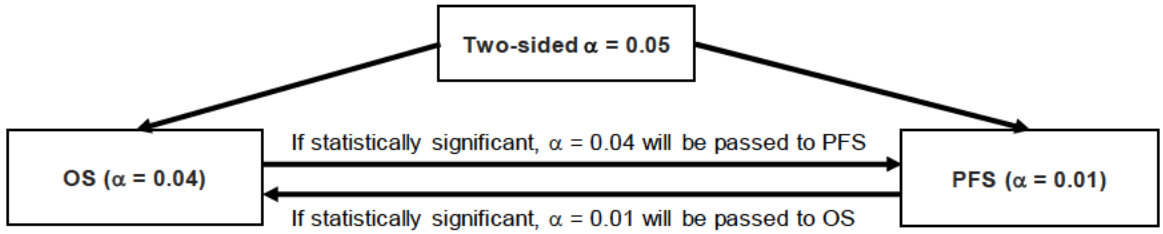
Approximately 450 patients are planned for enrollment.

3.1 TYPE I ERROR CONTROL

The type I error (α) for this study is 0.05 (two-sided) and will be controlled for the co-primary efficacy endpoints of OS and IRF-assessed PFS between treatment arms in the FAS.

To control the overall type I error rate at 0.05, a group sequential weighted Holm procedure (Ye et al. 2013) will be used wherein the two-sided α of 0.04 and 0.01 will be allocated to the primary endpoints of OS and PFS, respectively. If OS in the FAS is statistically significant at the two-sided α of 0.04, the α of 0.04 will be recycled to PFS in the FAS, which will then be tested at a two-sided α of 0.05. Additionally, if PFS in the FAS is statistically significant at the two-sided α of 0.01, the α of 0.01 will be recycled to OS in the FAS, which will then be tested at a two-sided α of 0.05. The study will be considered positive if statistical significance is achieved for either of the two co-primary endpoints. An overview of the type I error rate control strategy is shown in Figure 2.

Figure 2 Overview of the Type I Error Control for Co-Primary Endpoints



OS = overall survival; PFS = progression-free survival.

3.2 CO-PRIMARY ENDPOINT: OVERALL SURVIVAL

The sample size of the study was determined on the basis of the number of deaths (OS events) required in the FAS to demonstrate efficacy in terms of OS. To detect an improvement in OS through use of a log rank test at a two sided significance level of 0.04, approximately [redacted] OS events ([redacted] patients) will be required at the final OS analysis to achieve an overall [redacted] power, assuming a target HR of [redacted]. The minimum detectable difference (MDD) is an OS HR of [redacted] (median OS improvement of [redacted] months, from [redacted] months in the Placebo+PC arm to [redacted] months in the Atezo+Tira+PC arm). The final analysis of OS is expected to occur at approximately [redacted] months after the first patient is randomized. The calculation of sample size and estimates of the OS analysis timeline are based on the following assumptions:

- Patient randomization in a [redacted] ratio to the Atezo+Tira+PC arm or the Placebo+PC arm
- One-piece exponential distribution for OS in each arm
- Median OS of [redacted] months in the Placebo+PC arm and [redacted] months in the Atezo+Tira+PC arm (increase of [redacted] months, corresponding to a target HR of [redacted])
- Annual dropout rate of [redacted] for OS in each arm

- [redacted] planned OS interim analysis (IA) at approximately [redacted] of the information fraction, using the O'Brien-Fleming stopping boundaries approximated by the Lan-DeMets α-spending function (Demets et al. 1994)
- Recruitment of [redacted] patients will take place over approximately [redacted] months

3.3 CO-PRIMARY ENDPOINT: PROGRESSION-FREE SURVIVAL

The primary analysis of IRF-assessed PFS will take place when approximately [redacted] PFS events ([redacted] patients) or approximately [redacted] OS events ([redacted] patients) have been observed in the FAS, whichever occurs later. The [redacted] PFS events provides an overall [redacted] power to detect a target HR of [redacted] for PFS using a log-rank test at a two-sided significance level of [redacted]. The MDD based on [redacted] PFS events is a PFS HR of [redacted] (median PFS improvement of [redacted] months, from [redacted] months in the Placebo+PC arm to [redacted] months in the Atezo+Tira+PC arm). The primary analysis of PFS is expected to occur at approximately [redacted] months after the first patient is randomized. The estimates are based on the following assumptions:

- Patient randomization in a [redacted] ratio to the Atezo+Tira+PC arm or the Placebo+PC arm
- One-piece exponential distribution for PFS in each arm
- Median PFS of [redacted] months in the Placebo+PC arm and [redacted] months in the Atezo+Tira+PC arm (increase of [redacted] months, corresponding to a target HR of [redacted])
- Annual dropout rate of [redacted] for PFS in each arm
- Recruitment of [redacted] patients will take place over approximately [redacted] months

4. ANALYSIS SETS

The analysis sets are defined in Table 3:

Table 3 Analysis Sets

Analysis Set	Definition
Full analysis set (FAS)	All randomized patients regardless of whether they received any study treatment
Safety-evaluable set	All randomized patients who received any amount of study treatment
Pharmacokinetic (PK)-evaluable set	All randomized patients who received any amount of study treatment and who have at least one post-baseline PK sample available
Atezolizumab anti-drug antibody (ADA)-evaluable set	All randomized patients who received any amount of atezolizumab treatment and who have at least one post-baseline ADA sample available
Tiragolumab ADA-evaluable set	All randomized patients who received any amount of tiragolumab treatment and who have at least one post-baseline ADA sample available

5. STATISTICAL ANALYSES

The analyses described in this SAP will supersede those specified in protocol for the purposes of a regulatory filing.

5.1 GENERAL CONSIDERATION

Unless otherwise specified, efficacy analyses will be performed in the FAS. Patients will be grouped according to the treatment assigned at randomization by IxRS.

Safety analyses will be conducted in the safety-evaluable set, with patients grouped according to the actual treatment received. Specifically, a patient will be included in the Atezo+Tira+PC arm in the safety analyses if the patient receives any amount of tiragolumab and/or atezolizumab, regardless of the initial treatment assignment at randomization.

Baseline measurements are defined as the last available measurement obtained prior to the patient receiving the first dose of any component of study treatment, or prior to randomization for patients who were not treated, unless otherwise noted.

5.2 PATIENT DISPOSITION

Study enrollment and reasons for discontinuation from the study will be summarized by treatment arm for the FAS. Study treatment disposition and reasons for discontinuation from study treatment will be summarized for the safety-evaluable set.

5.3 PRIMARY ENDPOINTS ANALYSIS

5.3.1 Definition of Co-Primary Endpoints

The co-primary efficacy endpoints for this study are OS and IRF-assessed PFS in the FAS. The corresponding estimands are defined in Table 2.

An overview of the type I error rate control strategy is shown in Figure 2, wherein the two-sided α of 0.04 and 0.01 will be allocated to the primary endpoints of OS and PFS, respectively. If OS in the FAS is statistically significant at the two-sided α of 0.04, the α of 0.04 will be recycled to PFS in the FAS, which will then be tested at a two-sided α of 0.05. Additionally, if PFS in the FAS is statistically significant at the two-sided α of 0.01, the α of 0.01 will be recycled to OS in the FAS, which will then be tested at a two-sided α of 0.05.

5.3.1.1 Overall Survival

OS is defined as the time from randomization to death from any cause. Patients who are not reported as having died at the time of analysis will be censored at the last date they were known to be alive. Patients with no postbaseline survival information will be censored at the date of randomization.

5.3.1.2 IRF-Assessed Progression-Free Survival

IRF-assessed PFS is defined as the time from randomization to the first occurrence of disease progression or death from any cause (whichever occurs first), as determined by an IRF according to RECIST v1.1. Patients who have not experienced disease progression or death at the time of analysis will be censored at the time of the last tumor assessment. Patients with no postbaseline tumor assessment will be censored at the date of randomization.

5.3.2 Main Analytical Approach for Co-Primary Endpoints

The two-sided log-rank test, stratified by the protocol-defined stratification factors as entered in IxRS, will be used to compare OS and IRF-assessed PFS between the treatment arms. Cox proportional hazards model, stratified by the protocol-defined stratification factors as entered in IxRS, will be used to estimate the HR between the two treatment arms and its 95% confidence interval (CI).

Kaplan-Meier methodology will be used to estimate the median OS and PFS and to construct survival curves for each treatment arm for a visual description of the difference between treatment arms. The Brookmeyer-Crowley methodology will be used to construct the 95% CI for the median OS and PFS for each treatment arm (Brookmeyer and Crowley 1982).

5.3.3 Handling of Missing Data

5.3.3.1 Overall Survival

Patients who are lost to follow-up will be censored at the last date they were known to be alive for the primary analysis of OS. If >5% of patients are lost to follow-up for OS in either treatment arm, a sensitivity analysis will be performed for the comparisons between two treatment arms in which patients who are lost to follow-up will be considered as having died at the last date they were known to be alive.

5.3.3.2 IRF-Assessed Progression-Free Survival

The impact of missing scheduled tumor assessments on PFS will be assessed depending on the number of patients who missed tumor assessments scheduled immediately prior to the date of disease progression per RECIST v1.1 or the data cutoff. If >5% of patients missed two or more assessments scheduled immediately prior to the date of disease progression per RECIST v1.1 or the data cutoff in any treatment arm, the following two sensitivity analyses will be performed:

- Patients who missed two or more scheduled assessments immediately prior to the date of disease progression per RECIST v1.1 or the data cutoff will be censored at the last tumor assessment prior to the missed visits.
- The impact of missing scheduled tumor assessments on PFS will be assessed by performing a sensitivity analysis based on the interval censoring analysis method.

To apply the interval censoring, for each patient, the left and the right boundaries of the interval will be derived based on the following rules:

Table 4 Censoring Boundaries for PFS Sensitivity Analyses

Situations	Left Boundary[a]	Right Boundary
Patients who had disease progression prior to death	The date of the last assessment that showed a progression-free status	The date of the first assessment that showed disease progression
Patients who died without disease progression	The date of the last assessment that showed a progression-free status	Death date
Patients who did not die nor had disease progression	The date of the last assessment that showed a progression-free status	Not applicable (Missing)

PFS = progression-free survival

[a] For patients who did not have any post-baseline assessment with progression-free status, the left boundary is the date of randomization.

The PFS survival curves will be estimated using the nonparametric maximum likelihood estimator ([NPMLE]; Turnbull 1974) for each treatment arm. The median PFS of each treatment arm will be reported and its 95% CI will be constructed based on the Brookmeyer-Crowley method (Brookmeyer and Crowley 1982). Stratified log-rank test proposed by Sun (Sun 1996) will be used to compare PFS between the treatment arms. The treatment effect will be estimated using a stratified proportional hazard regression model (Finkelstein 1986) with a parametric assumption of piecewise exponential distribution for the baseline hazard function (Friedman 1982, Royston and Parmar 2002).

5.3.4 Sensitivity Analyses for Co-Primary Endpoints

Sensitivity analyses of the primary endpoints will be performed to assess the impact of stratification. These analyses will follow the same analyses method as the primary endpoints with the exception that OS and IRF-assessed PFS will be compared using an unstratified log-rank test and the HR will be estimated from unstratified Cox proportional hazards model (see also Section 5.3.2).

If at the time of the primary analysis for IRF-assessed PFS, the observed PFS events are more than [redacted] over the protocol-specified events of [redacted] a sensitivity analysis using an earlier data cutoff date with targeted [redacted] events will be performed.

5.3.5 Supplementary Analyses for Co-Primary Endpoints

The following supplementary analyses will be performed for the co-primary efficacy endpoints of OS and IRF-assessed PFS in which a different handling rule of intercurrent events is implemented to provide further understanding of the treatment effect.

To assess the impact of the intercurrent event of starting a non-protocol anti-cancer therapy (NPT) prior to a PFS event, the primary analysis of IRF-assessed PFS will be repeated with such intercurrent event handled using a hypothetical strategy, if >5% of patients received NPT prior to a PFS event in either treatment arm. To estimate the estimand that implements this strategy, patients who start an NPT before a PFS event will be censored at the time of the last tumor assessment before the initiation of NPT.

To assess the impact of the intercurrent event of starting an NPT on OS, the primary analysis of OS will be repeated with such intercurrent event handled using a hypothetical strategy. Rank-preserving structural failure time (RPSFT) method will be used to provide an estimate of the OS time for the treatment arms had NPT not been received (Robins and Tsiatis 1991). This method estimates OS measured from the time of NPT by applying an estimate of the benefit of the NPT. The total OS time (sum of time to NPT and the estimated survival time after NPT is started) will then be analyzed using the same methodology as for the primary analysis of OS.

5.3.5.1 Subgroup Analyses for Co-Primary Endpoints

The generalizability of OS and IRF-assessed PFS results when comparing the experimental arm to the control arm will be investigated by estimating the treatment effect in subgroups defined by demographics and baseline prognostic characteristics, including but not limited to: age, sex, race, ethnicity, ECOG performance status, alcohol use history, tobacco use history, PD-L1 expression by SP263, previous curative treatment, disease status, and number of organs with metastases.

Summaries of OS and IRF-assessed PFS, including unstratified HRs estimated from Cox proportional hazards models and Kaplan-Meier estimates of median OS and PFS, will be provided in forest plots.

5.4 SECONDARY ENDPOINTS ANALYSES

The estimands for selected secondary endpoints are defined in Table 2.

5.4.1 Investigator-Assessed Progression-Free Survival

Investigator-assessed PFS is defined as the time from randomization to the first occurrence of disease progression or death from any cause (whichever occurs first), as determined by the investigator according to RECIST v1.1. Patients who have not experienced disease progression or death at the time of analysis will be censored at the time of the last tumor assessment. Patients with no postbaseline tumor assessment will be censored at the date of randomization.

Methods for the investigator-assessed PFS analysis will be similar to those described for IRF-assessed PFS.

5.4.2 Confirmed Objective Response Rate

IRF-Assessed Objective Response Rate

An objective response per IRF is defined as a complete response (CR) or partial response (PR) as determined by an IRF according to RECIST v1.1. Patients not meeting these criteria, including patients without any postbaseline tumor assessment, will be considered non-responders. Confirmed objective response rate (ORR) is defined as the proportion of patients who achieved an objective response on two consecutive occasions ≥4 weeks apart. Patients must have measurable disease at baseline to be included in the analysis.

The two-sided Cochran-Mantel-Haenszel test, stratified by the same stratification factors used for the analyses of the co-primary endpoints, will be used to compare ORR between the two treatment arms. ORR will be calculated for each treatment arm, and the difference in ORR between treatment arms will be computed. The 95% CI for ORR for each arm will be derived through use of the Wilson score method. The 95% CI for difference in ORR will be computed using the Newcombe method.

Investigator-Assessed Objective Response Rate

ORR analyses will be performed separately on the basis of investigator-assessed tumor response according to RECIST v1.1. The analysis methods will be similar to those described for IRF-assessed ORR.

5.4.2.1 Duration of Response

IRF-Assessed Duration of Response

Duration of response (DOR) per IRF will be assessed in patients who achieved a confirmed objective response, as determined by an IRF according to RECIST v1.1. DOR is defined as the time from the first occurrence of a confirmed objective response (CR or PR, whichever status is recorded first) to the first occurrence of disease progression or death from any cause, whichever occurs first. Patients who have not progressed and who have not died at the time of analysis will be censored at the date of the last tumor assessment. The analysis of DOR is based on a non-randomized subset of patients (specifically, patients who achieved a confirmed objective response); therefore, formal hypothesis testing will not be performed for this endpoint. Comparisons between treatment arms will be made for descriptive purposes only.

Median DOR and corresponding 95% CIs will be estimated using Kaplan-Meier methodology for each treatment arm.

Investigator-Assessed Duration of Response

DOR analyses will be performed separately on the basis of investigator-assessed tumor response according to RECIST v1.1. The analysis methods will be similar to those described for IRF-assessed DOR.

5.4.2.2 Time to Confirmed Deterioration in Patient-Reported Outcomes

Time to confirmed deterioration (TTCD) in patient-reported physical functioning, role functioning, and global health status and quality of life (GHS/QoL), as measured by the respective scales of the European Organisation for Research and Treatment of Cancer (EORTC) Quality-of-Life Questionnaire Core 30 (QLQ-C30) questionnaire, will be defined as the time from randomization to first confirmed deterioration (decrease from baseline of ≥10 points) that is maintained for two consecutive assessments or followed by death from any cause within 3 weeks.

TTCD in patient-reported dysphagia, as measured by the dysphagia scale of EORTC Quality of Life Questionnaire - Oesophageal Cancer Module (QLQ-OES18) questionnaire, will be defined as the time from randomization to first confirmed deterioration (increase from baseline of ≥10 points) that is maintained for two consecutive assessments or followed by death from any cause within 3 weeks.

Patients who have not experienced a confirmed deterioration by the data cutoff date will be censored at the last time they complete an assessment. Patients with no baseline or post-baseline assessment will be censored at the date of randomization. According to the while-alive strategy, patients who die before reporting any confirmed deterioration will be censored at the last time they complete an assessment. The analysis methods will be similar to those described for OS.

[redacted]

5.6 SAFETY ANALYSES

Safety analyses will be conducted in the safety-evaluable set (see Section 4), with patients grouped according to the actual treatment received. Specifically, a patient will be included in the Atezo+Tira+PC arm in the safety analyses if the patient receives any amount of tiragolumab and/or atezolizumab, regardless of the initial treatment assignment at randomization.

5.6.1 Extent of Exposure

Study drug exposure, including treatment duration, number of cycles, and dose intensity, will be summarized by treatment arm for each study drug with descriptive statistics.

5.6.2 Adverse Events

Verbatim description of adverse events (AEs) will be mapped to MedDRA thesaurus terms. Severity for all events will be graded by the investigator according to National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE), Version 5.0, [redacted]

All treatment-emergent AEs will be summarized by treatment arm and NCI CTCAE grade. [redacted] will also be summarized by treatment arm and the [redacted] In addition, common AEs, treatment-related AEs, serious adverse events (SAEs), AEs leading to study treatment discontinuation, AEs leading to study treatment interruption or dose modification, Grade 3-4 AEs, and fatal AEs (Grade 5) will be summarized accordingly. For the purpose of analyses, AEs of special interest (AESI)

will be identified by a set of comprehensive definitions using standardized MedDRA queries, high-level terms, preferred terms, and Sponsor-defined AE grouped terms from the clinical database and presented by medical concepts. AESIs will be summarized by treatment arm and Common Terminology Criteria for Adverse Events (CTCAE) grade.

Multiple occurrences of the same event in a subject will be counted once at the highest grade.

For the safety analyses, “treatment-emergent” is defined as AEs occurring on or after the first dose of study drug treatment or as a pre-existing condition that worsened on or after the first dose of the study treatment up to the data cutoff date.

Listings of AEs will include all treatment emergent AEs up to the data cutoff date.

All deaths and causes of deaths will be summarized by treatment arm.

5.6.3 Laboratory Data

Laboratory data with values outside the normal ranges will be identified. Additionally, selected laboratory data will be graded according to NCI CTCAE v5.0 and will be summarized by treatment arm with shift tables from baseline to worst post baseline value.

5.6.4 Vital Signs

Changes in vital signs will be summarized.

5.7 OTHER ANALYSES

5.7.1 Summaries of Conduct of Study

Study enrollment and major protocol deviations including major deviations with regard to the inclusion and exclusion criteria will be summarized by treatment arm for the FAS.

5.7.2 Summaries of Treatment Group Comparability

Demographics (including age, sex, race) and baseline characteristics (e.g., PD-L1 expression, previous curative treatment, and ECOG Performance Status) will be summarized overall and by treatment arm. Descriptive statistics (mean, median, standard deviation, and range) will be presented for continuous data, and frequencies and percentages will be presented for categorical data, as appropriate.

5.7.3 Pharmacokinetic Analyses

Pharmacokinetic (PK) samples will be collected from all patients enrolled in the study.

Tiragolumab and atezolizumab serum concentration data (minimum serum concentration [C_{min}] and maximum serum concentration [C_{max}]) will be tabulated and summarized. Descriptive statistics will include mean, median, range, and standard deviation, as appropriate.

Additional PK analyses will be conducted, as appropriate, based on the availability of data.

5.7.4 Immunogenicity Analyses

The immunogenicity analyses will include patients with any anti-drug antibody (ADA) assessments, with patients grouped according to treatment received.

The numbers and proportions of treatment-emergent ADA-positive patients and ADA-negative patients during both the treatment and follow-up periods for both tiragolumab and atezolizumab will be summarized by treatment arm.

[redacted]

5.7.5 Health Status Utility Analyses

Health utility data from the EQ-5D-5L will be evaluated in pharmacoeconomic models. Results from health economic data analyses will be reported separately from the Clinical Study Report (CSR).

[redacted]

6. SUPPORTING DOCUMENTATION

This section is not applicable, since there is no additional supporting document.

7. REFERENCES

Brookmeyer R, Crowley J. A confidence interval for the median survival time. Biometrics 1982;38:29-41.

Cocks K, King MT, Velikova G, et al. Evidence-based guidelines for determination of sample size and interpretation of the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. J Clin Oncol 2011;29:89-96.

Demets DL, Lan KG. Interim analysis: the alpha spending function approach. Stat Med. 1994;13:1341-52.

Finkelstein DM. A proportional hazards model for interval-censored failure time data. Biometrics 1986:845-54.

Friedman, M. Piecewise exponential models for survival data with covariates. Ann Stat 1982;10:101-13.

Osoba D, Rodrigues G, Myles J, et al. Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol 1998;16:139-44.

Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Commun Stat Theory Methods 1991;20:2609-31.

Royston P, Parmar MK. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Stat Med 2002;21:2175-97.

Sun J. A non-parametric test for interval-censored failure time data with application to AIDS studies. Stat Med 1996;15:1387-95.

Turnbull BW. Nonparametric estimation of a survivorship function with doubly censored data. J Am Stat Assoc 1974;69:169-73.

Ye Y, Li A, Liu L, et al. A group sequential Holm procedure with multiple primary endpoints. Stat Med 2013;32:1112-24.