Official Title: A Phase III, Multicenter, Randomized, Open-Label Study to Evaluate the Efficacy and Safety of Atezolizumab Given in Combination With Cabozantinib Versus Cabozantinib Alone in Patients With Inoperable, Locally Advanced, or Metastatic Renal Cell Carcinoma who Experienced Radiographic Tumor Progression During or After Immune Checkpoint Inhibitor Treatment

NCT Number: NCT04338269

Document Date: SAP Amendment Version 2: 02-Sep-2022

STATISTICAL ANALYSIS PLAN

STUDY TITLE:	**A PHASE III, MULTICENTER, RANDOMIZED, OPEN-LABEL STUDY TO EVALUATE THE EFFICACY AND SAFETY OF ATEZOLIZUMAB GIVEN IN COMBINATION WITH CABOZANTINIB VERSUS CABOZANTINIB ALONE IN PATIENTS WITH INOPERABLE, LOCALLY ADVANCED, OR METASTATIC RENAL CELL CARCINOMA WHO EXPERIENCED RADIOGRAPHIC TUMOR PROGRESSION DURING OR AFTER IMMUNE CHECKPOINT INHIBITOR TREATMENT**
STUDY NUMBER:	WO41994
VERSION NUMBER:	2
ROCHE COMPOUND(S):	Atezolizumab (RO5541267)
EUDRACT NUMBER:	2020-000502-29160
IND NUMBER:	119,039
NCT NUMBER:	NCT04338269
PLAN PREPARED BY:	[redacted], M.S.

STATISTICAL ANALYSIS PLAN AMENDMENT APPROVAL

SPONSOR:	F. Hoffmann-La Roche Ltd
LEGAL REGISTERED ADDRESS:	Grenzacherstrasse 124 4070 Basel, Switzerland
DATE FINAL:	See electronic date stamp on the last page of this document

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)
2	see electronic date stamp on title page	Version 4, 14 February 2022
1	28 March 2022	Version 4, 14 February 2022

STATISTICAL ANALYSIS PLAN AMENDMENT RATIONALE

Key changes to the SAP, along with the rationale(s) for each change, are summarized below.

Section	Description of Change	Rationale for Change
5.3.3.1	Updated criteria for conducting the sensitivity analyses for PFS: removed the pre-specified criteria of >5% of patients in either treatment arm who missed two or more tumor assessments scheduled immediately prior to the date of disease progression or death in any treatment arm	Based on FDA feedback of SAP version 1, received on 11 August 2022

Additional minor changes have been made throughout to improve clarity and consistency.

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
CI	confidence interval
CR	complete response
CTCAE	Common Terminology Criteria for Adverse Events
DOR	duration of response
DRS-P	Disease-Related Symptom-Physical
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	electronic Case Report Form
EORTC	European Organization for Research and Treatment of Cancer
FKSI-19	Functional Assessment of Cancer Therapy–Kidney Symptom Index 19
GHS	global health status
HR	hazard ratio
ICI	immune checkpoint inhibitor
iDMC	independent Data Monitoring Committee
IMDC	International Metastatic Renal Cell Carcinoma Database Consortium
IRF	Independent Review Facility
ITT	intent-to-treat
IV	intravenous
IxRS	interactive voice/web-based response system
KPS	Karnofsky Performance Status
MDD	minimally detectable difference
MedDRA	Medical Dictionary for Regulatory Activities
NCI	National Cancer Institute
NPT	non-protocol anti-cancer therapy
OBF	O'Brien-Fleming
ORR	objective response rate
OS	overall survival
PD-L1	programmed death-ligand 1
PF	physical functioning
PFS	progression-free survival
PK	pharmacokinetic

Abbreviation or Term	Description
PR	partial response
PRO	patient-reported outcome
Q3W	every 3 weeks
QLQ-C30	Quality-of-Life Questionnaire Core 30
QD	once a day
QoL	Quality of Life
RCC	renal cell carcinoma
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	serious adverse events
SAP	Statistical Analysis Plan
VEGFR-TKI	vascular endothelial growth factor receptor-tyrosine kinase inhibitor

1. INTRODUCTION

This Statistical Analysis Plan (SAP) provides details of the planned analyses and statistical methods for Study WO41994, a Phase III, multicenter, randomized, open-label study of atezolizumab given in combination with cabozantinib versus cabozantinib alone in patients with inoperable, locally advanced, or metastatic renal cell carcinoma (RCC) who experienced radiographic tumor progression during or after immune checkpoint inhibitor (ICI) treatment in the adjuvant and/or locally advanced/metastatic setting. Detailed background information on the study can be found in the study protocol.

1.1 OBJECTIVES AND ENDPOINTS

Study WO41994 evaluates the efficacy and safety of atezolizumab in combination with cabozanitinib (herein referred to as Atezo+Cabo) compared with cabozanitinib alone (herein referred to as Cabo) in RCC patients with an immediate preceding treatment of ICI in the adjuvant and/or locally advanced/metastatic setting. Specific objectives and corresponding endpoints for the study are outlined in Table 1.

Table 1 Objectives and Corresponding Endpoints

Primary Efficacy Objective	Corresponding Endpoints
• To evaluate the efficacy of atezolizumab in combination with cabozantinib compared with cabozantinib alone	• Progression-free survival (PFS) assessed by an Independent Review Facility (IRF), defined as the time from randomization to the first occurrence of disease progression, as assessed by an IRF according to Response Evaluation Criteria in Solid Tumors, Version 1.1 (RECIST v1.1) or death from any cause, whichever occurs first • Overall survival (OS), defined as the time from randomization to death from any cause
Secondary Efficacy Objective	**Corresponding Endpoints**
• To evaluate the efficacy of atezolizumab in combination with cabozantinib compared with cabozantinib alone	• PFS assessed by the investigators, defined as the time from randomization to the first occurrence of disease progression, as assessed by the investigators according to RECIST v1.1 or death from any cause, whichever occurs first • Investigator- and IRF-assessed objective response rate (ORR), defined as the proportion of patients with a complete response or partial response on two consecutive occasions at least 4 weeks apart according to RECIST v1.1

	• Investigator- and IRF-assessed duration of response, defined as the time from the first occurrence of a documented, confirmed objective response to disease progression according to RECIST v1.1, or death from any cause, whichever occurs first
Safety Objective	**Corresponding Endpoints**
• To evaluate the safety of atezolizumab in combination with cabozantinib compared with cabozantinib	• Incidence and severity of adverse events, with severity determined according to National Cancer Institute Common Terminology Criteria for Adverse Events, Version 5.0 • Change from baseline in targeted vital signs • Change from baseline in targeted clinical laboratory test results
Pharmacokinetic Objective	**Corresponding Endpoints**
• To characterize the pharmacokinetic (PK) profile of atezolizumab and cabozantinib	• Serum concentrations of atezolizumab at specified timepoints • Plasma concentrations of cabozantinib at specified timepoints
Immunogenicity Objective	**Corresponding Endpoints**
• To evaluate the immune response to atezolizumab	• Prevalence of anti-drug antibodies (ADAs) to atezolizumab at baseline and incidence of ADAs to atezolizumab during the study • Relationship between ADA status and demographics, efficacy, safety, or PK endpoints
Exploratory Objectives	**Corresponding Endpoints**
To evaluate the efficacy of atezolizumab in combination with cabozantinib compared with cabozantinib alone	• PFS, OS, and ORR in subgroups, defined by demographic and baseline characteristics • Time to response
To evaluate the patient-reported outcomes of atezolizumab in combination with cabozantinib compared with cabozantinib alone	• Time to confirmed deterioration in symptoms based on the Functional Assessment of Cancer Therapy–Kidney Symptom Index 19 (FKSI-19) Disease-Related Symptom-Physical (DRS-P) scale • Time to confirmed deterioration in physical functioning (PF) and global health status/quality of life (GHS/QoL), based on the corresponding scales from the European Organisation for Research and Treatment of

	Cancer (EORTC) Quality of Life Questionnaire-Core 30 (QLQ-C30) • Change from baseline in symptoms, function, and QoL, based on FKSI-19 DRS-P scale, EORTC QLQ-C30 PF scale, and EORTC QLQ-C30 GHS/QoL scale by visit • Overall bother with treatment side effects during study treatment, based on FKSI-19 GP5 item • Change from baseline in the EQ-5D-5L index-based and visual analog scale scores by visit
• To identify and/or evaluate biomarkers	• Relationship between biomarkers in tumor tissue and blood and efficacy, safety, PK, or other biomarker endpoints

ADA = anti-drug antibody; DRS-P = Disease-Related Symptom-Physical; EORTC = European Organisation for the Research and Treatment of Cancer; FKSI-19 = Functional Assessment of Cancer Therapy–Kidney Symptom Index 19; GHS = Global Health Status; IRF= Independent Review Facility; NCI = National Cancer Institute; ORR = objective response rate; OS = overall survival; PFS = progression-free survival; PK = pharmacokinetic; QLQ-C30=Quality-of-Life Questionnaire Core 30; QoL = Quality of Life; RECIST = Response Evaluation Criteria in Solid Tumors.

1.2 STUDY DESIGN

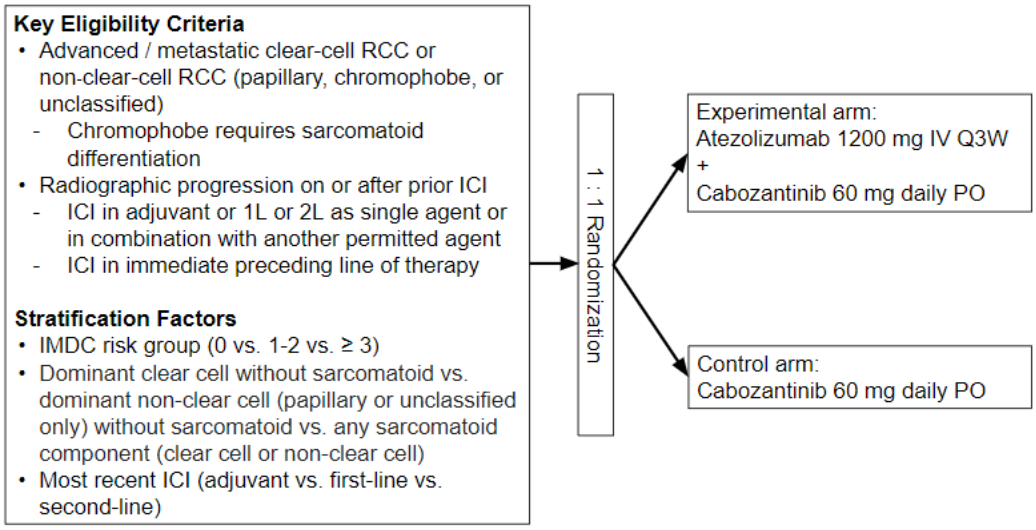
This is a Phase III, multicenter, randomized, open-label study designed to evaluate the efficacy and safety of atezolizumab given in combination with cabozantinib versus cabozantinib alone in patients with inoperable, locally advanced, or metastatic RCC who experienced radiographic tumor progression during or after ICI treatment in the adjuvant and/or locally advanced/metastatic setting. The study will enroll approximately 500 patients at approximately 140-180 sites globally.

Eligible patients will be randomized in a 1:1 ratio to one of the following two treatments arms:

- Experimental arm (Atezo+Cabo): Atezolizumab 1200 mg intravenous (IV) infusions every 3 weeks (Q3W) on Day 1 of each 21-day cycle plus cabozantinib 60-mg oral tablets taken once a day (QD; 1 cycle=21 days)
- Control arm (Cabo): Cabozantinib 60-mg oral tablets taken QD (1 cycle=21 days)

The study schema is shown in Figure 1.

Figure 1 Study Schema



1L = first-line (treatment); 2L = second-line (treatment); ICI = immune checkpoint inhibitor; IMDC = International Metastatic Renal Cell Carcinoma Database Consortium; PO = by mouth; Q3W = every 3 weeks; RCC = renal cell carcinoma.

Atezolizumab will be administered by IV infusion at a fixed dose of 1200 mg on Day 1 of each 21-day cycle and cabozantinib will be taken orally at a starting dose of 60 mg/day on Days 1–21 of each 21-day cycle. Patients randomized to the Atezo+Cabo arm who transiently withhold or permanently discontinue either atezolizumab or cabozantinib may continue on single-agent therapy until disease progression (i.e., patients being withheld from cabozantinib transiently for adverse effects may continue atezolizumab monotherapy and vice versa).

Patients will receive atezolizumab and/or cabozantinib until disease progression per RECIST v1.1, unacceptable toxicity, or symptomatic deterioration attributed to disease progression (e.g., pain secondary to disease or unmanageable ascites) as determined by the investigator after an integrated assessment of radiographic and biochemical data, local biopsy results (if available), and clinical status.

No crossover will be allowed from the control arm to the experimental arm.

Patients will undergo scheduled tumor assessments at baseline, every 9 weeks (±7 days) for the first 18 months, and every 12 weeks (±7 days) thereafter. Tumor assessments will continue until disease progression as assessed by the investigator per RECIST v1.1 or, for patients who continue study treatment after radiographic disease progression, loss of clinical benefit as determined by the investigator. In the absence of disease progression, tumor assessments should continue regardless of whether treatment has been discontinued (e.g., for toxicity) or whether patients start new anti-cancer therapy, until consent is withdrawn, death, or the study is terminated by the Sponsor, whichever occurs first.

All patients, regardless of arm, are required to perform a follow-up post-disease progression tumor assessment after investigator assessment of radiographic disease progression per RECIST v1.1. This subsequent scan will take place on the same schedule as prior to progression at 9 weeks (± 7 days) if disease progression occurred in the first 18 months, and at 12 weeks (± 7 days) if disease progression occurred after the first 18 months following treatment initiation.

Following treatment discontinuation, patients will be followed for survival and subsequent anti-cancer therapies until death, loss to follow-up, withdrawal of consent, or study termination by Sponsor, whichever occurs first.

1.2.1 Treatment Assignment and Blinding

This is a randomized open-label 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 1 of 2 treatment arms: Atezo+Cabo arm or Cabo arm. Randomization will occur in a 1:1 ratio with 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:

- International Metastatic Renal Cell Carcinoma Database Consortium (IMDC) risk groups (favorable, intermediate, or poor risk; 0, 1–2, or ≥3), which comprises the following 6 risks factors: time from diagnosis to systemic therapy, Karnofsky Peformance Status (KPS), hemoglobin, corrected calcium, neutrophil, and platelet count
- Most recent ICI therapy (adjuvant vs. locally advanced/metastatic first-line vs. locally advanced/metastatic second-line)
- Histology: dominant clear-cell without sarcomatoid versus dominant non–clear-cell (papillary or unclassified only) without sarcomatoid versus any sarcomatoid component (with clear-cell or non–clear-cell)

1.2.2 Independent Review Facility

An idependent review facility (IRF) will be used to conduct blinded radiology review of the imaging data and will provide an independent assessment of tumor response and progression for all patients. Independent Review Facility-assessed endpoints will be used for primary and secondary analyses.

All scans performed as part of defined tumor assessments must be submitted to an IRF for central review.

1.2.3 Data Monitoring

An independent Data Monitoring Committee (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.

Safety data will be reviewed on a periodic basis, approximately every 6 months from the time of enrollment of the first patient until the time of the analysis of the primary efficacy endpoint of PFS according to policies and procedures detailed in an iDMC Charter. No interim efficacy analyses are planned for PFS.

All summaries and analyses for the iDMC review will be prepared by an independent Data Coordinating Center. The safety summaries will include demographic data, adverse events (AEs), serious adverse events (SAEs), and relevant laboratory data.

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/Ethics Committees.

2. STATISTICAL HYPOTHESES

The purpose of this study is hypothesis testing and estimation regarding the effect of atezolizumab in combination with cabozantinib on the duration of progression-free survival (PFS) and/or overall survival (OS) compared with cabozantinib alone.

The null hypothesis of no difference in PFS or OS between the two treatment arms in the intent-to-treat (ITT) population will be tested using the stratified log-rank test. The null and alternative hypotheses in terms of the survival functions $S_A(t)$ and $S_B(t)$ in Atezo+Cabo and Cabo arms, respectively, are phrased as below:

$$H_0: S_A(t) = S_B(t) \text{ versus } H_1: S_A(t) \neq S_B(t)$$

The Hazard Ratio (HR), λ_A/λ_B, where λ_A and λ_B represent the hazard rates for having a PFS or OS event in the Atezo+Cabo and Cabo arms, respectively, and the respective 95% CI will be estimated using the proportional Cox regression model.

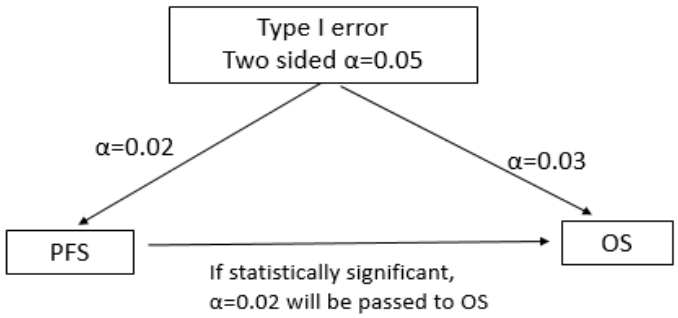
3. SAMPLE SIZE DETERMINATION

Approximately 500 patients are planned for enrollment globally over 20 months. The sample size calculation is determined based on the below considerations.

3.1 TYPE I ERROR CONTROL

The type I error (α) for the entire study is 0.05 (2-sided; Figure 2). There are multiple primary efficacy endpoints for this study: PFS by IRF assessment (IRF-PFS) per RECIST v1.1 and OS in the ITT population. To control the overall type I error rate (Bretz et al. 2009) at $\alpha = 0.05$ while accounting for 2 primary endpoints, α is split between PFS ($\alpha = 0.02$) and OS ($\alpha = 0.03$). The type I error can be recycled (Burman et al. 2009) if PFS results in the ITT population are statistically significant at $\alpha = 0.02$, then $\alpha = 0.02$ will be recycled to OS in the ITT population, and OS in the ITT population will be evaluated at $\alpha = 0.05$. The study will be considered as a positive study if statistical significance is achieved in favor of the experimental arm for either of the multiple primary endpoints, since the α for the entire study is controlled at 0.05.

Figure 2 Progression Free Survival and Overall Survival Analysis α-Allocation and α-Recycling



OS = overall survival; PFS = progression-free survival

3.2 PRIMARY ENDPOINT: IRF-PFS PER RECIST V1.1 IN THE ITT POPULATION

The analysis of the primary endpoint of IRF-PFS per RECIST v1.1 will take place when approximately 325 IRF-PFS events have occurred in the ITT population (65% events to patients ratio) based on the following assumptions:

- Two-sided, stratified log-rank test
- Patients randomized to Atezo+Cabo and Cabo arms in a 1:1 ratio
- PFS follows an exponential distribution
- $\alpha = 0.02$ (2-sided)
- Approximately 90% power
- Median PFS for the Cabo arm of 8.0 months and estimated median PFS in the Atezo+Cabo arm of 11.9 months (corresponding to HR of 0.67)
- 5% annual loss to follow-up for PFS
- No interim analysis

On the basis of these assumptions, it is projected that an observed HR of 0.77 or lower will result in a statistically significant difference between treatment arms (i.e., an HR of 0.77 will be the minimum detectable difference [MDD] for the analysis; this corresponds to an improvement of 2.4 months in median PFS from 8.0 months in the Cabo arm to 10.4 months in the Atezo+Cabo arm).

3.3 PRIMARY ENDPOINT: OS IN THE ITT POPULATION

The final analysis of the primary endpoint of OS will take place when approximately 325 OS events have occurred in the ITT population (65% events to patients ratio) based on the following assumptions:

- Two-sided, stratified log-rank test
- Patients randomized to Atezo+Cabo and Cabo arms in a 1:1 ratio
- OS follows an exponential distribution
- $\alpha = 0.03$ (2-sided)
- Approximately 85% power
- Median OS in the Cabo arm of 22 months and estimated median OS in the Atezo+Cabo arm of 31.4 months (an increase of 9.4 months, corresponding to an HR of 0.70)
- 1% annual loss to follow-up for OS
- Two interim OS analyses (see Section 5.8 for details)

At the final OS analysis, on the basis of these assumptions, it is projected that an observed OS HR of 0.78 or lower in the ITT population will result in a statistically significant difference between treatment arms (i.e., the MDD at the analysis; this

corresponds to an improvement of 6.2 months in median OS, from 22 months in the Cabo arm to 28.2 months in the Atezo+Cabo arm).

3.4 SAMPLE SIZE

With the above assumptions on IRF-PFS and OS, the sample size is determined at approximately 500 patients, where the IRF-PFS and OS final analyses will be conducted with sufficient statistical power for testing the target HRs or lower when approximately 325 events occur (65% events to patients ratio), respectively.

4. ANALYSIS SETS

The analysis sets used for analyses are defined in Table 2 below.

Table 2 Analysis Sets

Analysis set	Definition
ITT	All randomized patients, whether or not the patient received the assigned treatment
ORR-evaluable	All randomized patients with measurable disease at baseline
Safety-evaluable	All randomized patients who received at least one dose of study treatment
PK-evaluable	All patients who received at least one dose of study treatment and who have at least one post-baseline PK sample available
Atezolizumab ADA-evaluable	All patients who received at least one dose of atezolizumab treatment and with an ADA assay result from at least one sample result

ADA = anti-drug antibody; ITT = intent-to-treat; ORR = objective response rate
PK=pharmacokinetic

5. STATISTICAL ANALYSES

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

5.1 GENERAL CONSIDERATIONS

All efficacy analyses will be performed in the ITT poulation, unless otherwise specified. Patients will be analyzed according to the treatment assigned at randomization by IxRS, regardless of whether they receive any assigned study treatment.

Safety analyses will be conducted on the safety-evaluable patients, and will be performed based on the actual treatment patients received, regardless of the initial treatment assignment at randomization. Specifically, a patient will be included in the Atezo+Cabo arm in the safety analyses if the patient receives any amount of atezolizumab, regardless of the initial treatment assignment at randomization. Similarly, a patient will be included in the Cabo arm in the safety analyses if the patient receives

any amount of cabozanitib without any atezolizumab, regardless of the initial treatment assignment at randomization.

Unless otherwise stated, baseline values are the last available data obtained prior to the patient receiving the first dose of study treatment on Cycle 1, Day 1 (or at screening, for patients who were not treated).

5.2 PATIENT DISPOSITION

Study enrollment and reasons for discontinuation from the study will be summarized by assigned treatment arm for the ITT population. Study treatment disposition and reasons for discontinuation from study treatment will be summarized for the safety-evaluable patients by actual treatment arm.

5.3 PRIMARY ENDPOINT ANALYSES

5.3.1 Definition of Primary Endpoints

The multiple primary efficacy endpoints are IRF-PFS per RECIST v1.1 and OS.

IRF-PFS is defined as the time from randomization to disease progression, as determined by the IRF per RECIST v1.1, or death from any cause, whichever occurs first. Data for patients who have not experienced disease progression or death will be censored at the last tumor assessment date. Data for patients with no post-baseline tumor assessments will be censored at the randomization date.

Overall survival is defined as the time from randomization to death due to any cause. Data for patients who are not reported as having died at the date of analysis will be censored at the date when they were last known to be alive. Patients who do not have post-baseline information will be censored at the date of randomization.

The hypothesis testing for PFS in the ITT population will be conducted at a two-sided α of 0.02. If PFS is not statistically significant in the ITT population, the hypothesis testing for OS in the ITT population will be conducted at a two-sided α of 0.03. If PFS is statistically significant in the ITT population, OS will be tested at a two-sided α level of 0.05.

5.3.2 Main Analytical Approach for Primary Endpoints

The stratified log-rank test at the 2-sided 5% level of significance will be used to compare PFS and OS between the treatment arms, according to the protocol-defined stratification factors as entered in IxRS for the ITT patients. The stratified Cox proportional hazards model will be used to estimate the HRs between the two treatment arms and its 95% confidence interval (CI).

For the stratification factors, most recent ICI therapy (adjuvant vs. locally advanced/metastatic first-line vs. locally advanced/metastatic second-line), histology (dominant clear-cell without sarcomatoid vs. dominant non–clear-cell [papillary or

unclassified only] without sarcomatoid vs. any sarcomatoid component [clear-cell or non−clear-cell]), and the IMDC score (0, 1−2, ≥3), if at least one stratum has less than 10 events at the time of analysis, the stratification factor that contains the level with the smallest number of patients will be removed from the stratified analyses. The final set of stratification factors used for the multiple primary endpoints will be applied to all other endpoints where stratified analyses are planned. The stratification factors will be obtained from the IxRS at the time of randomization.

Results from an unstratified analysis for log-rank test and Cox proportional hazards model, respectively, will also be provided.

Kaplan-Meier methodology will be used to estimate the median PFS and OS for each treatment arm, and Kaplan-Meier curves will be produced. The Brookmeyer Crowley methodology will be used to construct the 95% CI for the median PFS and OS for each treatment arm (Brookmeyer and Crowley 1982).

5.3.3 Sensitivity Analyses for Primary Endpoints

5.3.3.1 Analysis of IRF-PFS Accounting for Missing Tumor Assessments

The impact of missing scheduled tumor assessments on IRF-PFS will be assessed with the following two sensitivity analyses:

- Patients who missed two or more consecutive scheduled tumor assessments immediately prior to the date of disease progression by IRF-assessment per RECIST v1.1 or death will be censored at the last tumor assessment prior to the missed visits.
- Patients who missed two or more consecutive scheduled tumor assessment immediately prior to the date of disease progression by IRF-assessment per RECIST v1.1 or death will be counted as having progressed on the date of the first of these missing tumor assessments

5.3.3.2 Analysis of OS Accounting for Loss to Follow-up

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 comparison 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.

5.3.3.3 Analysis of IRF-PFS and OS Accounting for Non-Protocol Anti-Cancer Therapy

The impact of non-protocol anti-cancer treatment (NPT) prior to an IRF-PFS or OS event will be assessed depending on the number of patients who receive NPT.

If >5% of patients received any NPT prior to an IRF-PFS or an OS event in either treatment arm, sensitivity analyses for IRF-PFS and OS will be performed for the comparisons between treatment arms in which patients who received NPT before the event will be censored at the last tumor assessment date before the initiation of NPT.

Additionally, if >5% of patients received any NPT prior to an IRF-PFS, a sensitivity analysis for IRF-PFS will be performed in which patients who received NPT prior to the IRF-PFS event will be considered as having progressed at the date of initiation of NPT.

5.3.3.4 Analysis of IRF-PFS and OS Accounting for Discrepancy of Stratification Factors Between IxRS and eCRF

The sensitivity analyses for IRF-PFS and OS based on the strata recorded in the electronic Case Report Forms (eCRF) will be performed for the comparsion between the treatment arms if >10% of patients have been assigned to a stratum by IxRS that is different from the one recorded in the eCRF.

Additional sensitivity analyses may be conducted.

5.3.4 Supplementary Analyses for Primary Endpoints

5.3.4.1 Subgroup Analyses for Primary Endpoints

The generalizability of IRF-PFS and OS results when comparing the Atezo+Cabo arm to the Cabo arm will be investigated by estimating the treatment effect across subgroups in the ITT population. The subgroups are defined by the following:

- Demographics (e.g., age, sex, and race/ethnicity)
- Baseline prognostic characteristics (e.g., programmed death-ligand 1 [PD-L1] status, prior vascular endothelial growth factor receptor-tyrosine kinase inhibitor [VEGFR-TKI] use, most recent ICI therapy [adjuvant vs. locally advanced/metastatic first-line treatment vs. locally advanced/metastatic second-line treatment], tumor histology, or IMDC risk group).

Summaries of IRF-PFS and OS, including unstratified HRs estimated from Cox proportional hazards models and Kaplan-Meier estimates of median IRF-PFS and OS, will be provided separately for each level of the subgroups for the comparisons between the treatment arms.

5.4 SECONDARY ENDPOINT ANALYSES

5.4.1 IRF-assessed Objective Response Rate

An IRF-assessed objective response is defined as either a complete response (CR) or partial response (PR; confirmation is required, i.e., with CR or PR at two consecutive

tumor assessments at least 28 days apart) based on RECIST v1.1, as assessed by an IRF. Patients not meeting this criterion, including patients without any post-baseline tumor assessments, will be considered non-responders. Objective response rate (ORR) is defined as the proportion of patients who had an objective response in the ORR-evaluable population, defined as patients with measurable disease at baseline. Unconfirmed response rate will also be evaluated.

Objective response rate will be compared between treatment arms with use of the stratified Cochran-Mantel-Haenszel test. The stratification factors will be the same as those described in the analysis of the multiple primary efficacy endpoints of IRF-PFS and OS. An estimate of ORR will be calculated for each treatment arm, and its 95% CI will be calculated with use of the Clopper-Pearson method. The difference in ORR between treatment arms will be calculated, and its 95% CI will be calculated with use of the normal approximation to the binomial distribution.

5.4.2 IRF-assessed Duration of Response

IRF-assessed duration of response (DOR) is defined for patients who had a confirmed objective response as the time from the first occurrence of response (CR or PR) to disease progression, as assessed by an IRF or death, whichever occurs first. Data for patients who have not experienced disease progression or death will be censored at the last tumor assessment date. If no tumor assessments were performed after the date of the first occurrence of CR or PR, data for DOR will be censored at the date of the first occurrence of CR or PR.

Duration of response is based on a non-randomized subset of patients (those who achieved an objective response); therefore, formal hypothesis testing will not be performed for this endpoint. Comparisons between two arms will be made for descriptive purposes only. Methods for comparison of DOR between treatment arms will be the same as the methods for treatment comparison for the multiple primary efficacy endpoints of IRF-PFS and OS.

5.4.3 Investigator-assessed PFS, ORR, and DOR

Investigator-assessed PFS, ORR, and DOR follow the same definitions and analysis methods for IRF-assessed ones with the exception that tumor assessments are performed by the investigators.

5.5 EXPLORATORY ENDPOINT ANALYSES

5.5.1 PFS and OS Rates at Selected Time Points

The PFS rates at selected time points (e.g., 6 and 12 months) are defined as the proportion of patients who have not experienced disease progression according to RECIST v1.1 or death from any cause at selected time points after randomization.

The OS rates at selected time points (e.g., 12 and 24 months) are defined as the proportion of patients who have not died from any cause at selected time points after randomization.

The IRF- or investigator-assessed PFS and OS rates at selected time points will be estimated by the Kaplan-Meier methodology for each treatment arm and the 95% CI will be calculated with use of Greenwood's formula.

5.5.2 Time to Response

Time to response is defined as the time from randomization to first response of PR or CR among responders, as assessed by the investigators and an IRF according to RECIST v1.1. Descriptive statistics will be used to summarize the mean, median, minimum, and maximum time to response.

5.5.3 Patient Reported Outcome

5.5.3.1 Time to Confirmed Deterioration in Disease-Related Symptoms

Time to confirmed deterioration of disease-related symptoms is defined as the time from randomization date to the date of a patient's first 4-point or more score decrease from baseline on the Functional Assessment of Cancer Therapy–Kidney Symptom Index 19 (FKSI-19) Disease-Related Symptom-Physical (DRS-P) scale held for at least two consecutive timepoints or followed by death within 3 weeks (if Cycles 1–12) or 6 weeks (if after Cycle 12) from the last patient reported outcome (PRO) assessment. Kaplan-Meier methods will be applied to this endpoint. Patients who have not experienced confirmed deterioration will be censored at the last time of completed assessment. Patients with no post-baseline assessments will be censored at the randomization date.

5.5.3.2 Time to Confirmed Deterioration in Physical Functioning

Time to confirmed deterioration in physical functioning (PF) is defined as the time from randomization date to the date of a patient's first 10-point or more score decrease from baseline on the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire-Core 30 (QLQ-C30) PF scale held for at least 2 consecutive timepoints or followed by death within 3 weeks (if Cycles 1–12) or 6 weeks (if after Cycle 12) from the last PRO assessment. Kaplan-Meier methods will be applied to this endpoint.

5.5.3.3 Time to Confirmed Deterioration in Global Health Status/Quality of Life

Time to confirmed deterioration in global health status/quality of life (GHS/QoL) is defined as the time from randomization date to the date of a patient's first 10-point or more score decrease from baseline on the EORTC QLQ-C30 GHS/QoL scale held for at least two consecutive timepoints or followed by death within 3 weeks (if Cycles 1–12) or

6 weeks (if after Cycle 12) from the last PRO assessment. Kaplan-Meier methods will be applied to this endpoint.

5.5.3.4 Patient-Reported Outcome Descriptive Summaries

Compliance rates in the ITT population will be calculated as the number of patients who completed the assessment divided by the number of patients expected to complete the assessment at each timepoint for each treatment arm. Reasons for missing assessments, if available, will be summarized with use of frequencies and percentages.

Descriptive analyses will include summary statistics (mean, standard deviation, median, interquartile range, minimum, maximum) of PRO scores and score changes from baseline at each assessment timepoint by treatment arm. Additional timepoints of interest include PRO score at radiographic disease progression (i.e., a patient's last PRO assessment score within 30 days prior to or on the day of diagnosis of disease progression) and treatment discontinuation due to adverse events (i.e., a patient's last PRO assessment within the 30 days prior to treatment discontinuation due to adverse events). Graphs of mean scores and/or score changes from baseline along with 95% CIs may be presented. Descriptive summaries will be reported for the key scales (FKSI-19 DRS-P, EORTC QLQ-C30 PF, and EORTC QLQ-C30 GHS/QoL) as well as the remaining FKSI-19 and EORTC QLQ-C30 scales. Linearly transformed scores for the EORTC QLQ-C30 scales (per the EORTC scoring manual) will be calculated.

Cumulative distribution function plots of score change from baseline to Month 6 by treatment arm will be presented for each key scale (FKSI-19 DRS-P, EORTC QLQ-C30 physical function, EORTC QLQ-C30 GHS/QoL).

These descriptive summaries may not be included in the Clinical Study Report for this study.

5.5.3.5 Patient-Reported Outcome Side-Effect Burden

Descriptive analysis of the patient-reported overall side-effect bother item (FKSI-19 GP5) will be performed by treatment arm at each visit in the safety-evaluable population. Distribution of responses will be summarized as frequencies and percentages. Change from baseline may be summarized as no change; improved by 1, 2, 3, or 4 levels; and worsened by 1, 2, 3, or 4 levels. Stacked bar charts may also be used to illustrate the distribution of responses or the change from baseline at each timepoint by treatment arm.

These analyses may not be included in the Clinical Study Report for this study.

5.5.3.6 Health Status Utility Analyses

To evaluate health status utility scores of patients treated with atezolizumab in combination with cabozantinib compared with cabozantinib alone, change from baseline in EQ-5D-5L health utility index-based, and visual analog scale scores will be calculated

at specified timepoints. EQ-5D-5L is collected for use in economic models for reimbursement.

These analyses will not be included in the Clinical Study Report for this study.

5.6 SAFETY ANALYSES

Unless specified otherwise, safety analyses described below will be conducted for the safety evaluable patients, with patients grouped according to actual treatment received as defined in Section 5.1 for safety analyses.

5.6.1 Extent of Exposure

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

5.6.2 Adverse Events

Verbatim description of AEs will be mapped to the Medical Dictionary for Regulatory Activities (MedDRA) thesaurus terms. Severity for all AEs will be graded by the investigator according to the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE), Version 5.0 (v5.0).

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

Incidence and severity of all treatment-emergent AEs will be summarized by treatment arm and NCI CTCAE grade. In addition, common AEs, treatment-related AEs, SAEs, AEs leading to study treatment discontinuation or interruption, Grade 3-4 AEs, and fatal AEs (Grade 5) will be summarized accordingly. For the purpose of analyses, adverse events of special interest, identified by a set of comprehensive definitions using standardized MedDRA queries, High-Level Terms, and Sponsor-defined adverse event grouped terms from the AE clinical database by medical concept, will be summarized by treatment arm and CTCAE grade. Medical concepts include atezolizumab-associated identified risks, potential risks, and class effects reported with other ICIs.

Multiple occurrence of the same event will be counted once at the maximum severity.

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

Deaths during the study treatment period and those reported during the follow-up period after treatment completion or discontinuation and causes of death will be summarized by treatment arm.

5.6.3 Laboratory Data

Laboratory data will be summarized by treatment arm. Selected laboratory data will be graded according to NCI CTCAE v5.0 and will be summarized descriptively. Shift tables from baseline to worst post-baseline values will also be presented.

5.6.4 Vital Signs

Vital signs, including diastolic and systolic blood pressure, pulse rate, respiratory rate, and temperature, outside of normal limits will be summarized by treatment arm.

5.6.5 ECGs

Electrocardiograms (ECGs) of clinical significant abnormality will be summarized by treatment arm.

5.7 OTHER ANALYSES

5.7.1 Summaries of Conduct of Study

Study enrollment and major protocol deviations, including major deviations of inclusion/exclusion criteria, will be summarized by treatment arm for the ITT population.

5.7.2 Summaries of Treatment Group Comparability

Demographic variables, such as age, sex, race/ethnicity, stratification factors (IMDC score, line of therapy, histology), and baseline characteristics will be summarized by treatment arm as well as for all patients in the ITT population. Continuous variables will be summarized with use of means, standard deviations, medians, and ranges. Categorical variables will be summarized by proportions.

Medical history will be summarized for the safety evaluable patients. The summary will be separately reported for resolved events or conditions versus ongoing events or conditions as collected at baseline. Concomitant medications, used by the patient within 7 days prior to initiation of study treatment, will also be summarized for the safety evaluable patients by treatment arm for medications taken prior to the first dose of study treatment regardless of whether medications were ongoing or not after starting treatment versus the initial medications taken after the first dose of study treatment.

5.7.3 Pharmacokinetic Analyses

Pharmacokinetic analyses will be conducted for the PK-evaluable population that consist of all patients who received at least one dose of study treatment and who have at least one post-baseline PK sample available.

Atezolizumab serum concentration data (minimum serum concentration [C_{min}] and maximum serum concentration [C_{max}]) will be tabulated and summarized. Descriptive statistics will include means, medians, ranges, coefficients of variation, and standard deviations, as appropriate.

Cabozantinib plasma concentration data will be summarized using descriptive statistics as described above.

The concentration-time course will also be plotted for each drug. Additional PK analyses may be conducted, as appropriate, based on the availability of the data.

5.7.4 Immunogenicity Analyses

The immunogenicity analyses will be conducted for the atezolizumab ADA-evaluable population that consist of all patients who received any amount of atezolizumab with at least one ADA assessment for atezolizumab.

The numbers and proportions of ADA-positive patients and ADA-negative patients at baseline (baseline prevalence) and after drug administration (post-baseline incidence) will be summarized. When determining post-baseline incidence, patients are considered to be ADA-positive if they are ADA-negative or have missing data at baseline but develop an ADA response following study drug exposure (treatment-induced ADA response), or if they are ADA-positive at baseline and the titer of one or more post-baseline samples is at least 0.60 titer unit greater than the titer of the baseline sample (treatment-enhanced ADA response). Patients are considered to be ADA-negative if they are ADA-negative or have missing data at baseline and all post-baseline samples are negative, or if they are ADA-positive at baseline but do not have any post-baseline samples with a titer that is at least 0.60 titer unit greater than the titer of the baseline sample (treatment unaffected).

The relationship between ADA status and safety, efficacy, PK, and biomarker endpoints may be analyzed and reported via descriptive statistics.

5.7.5 Biomarker Analyses

Exploratory biomarker analyses may be performed in an effort to understand the association of biomarkers in tumor tissue, blood, and urine with study treatment, and to identify and/or evaluate biomarkers that are predictive of response to atezolizumab in combination with cabozantinib or cabozantinib alone (i.e., predictive biomarkers), are early surrogates of efficacy, are associated with progression to a more severe disease state (i.e., prognostic biomarkers), are associated with acquired resistance to atezolizumab in combination with cabozantinib or cabozantinib alone, can provide evidence of atezolizumab in combination with cabozantinib or cabozantinib alone efficacy (i.e., pharmacodynamic biomarkers), or can increase the knowledge and understanding of disease biology and drug safety.

These exploratory analyses may not be included in the Clinical Study Report for this study.

5.7.6 Analyses of Subgroups of Interest

The consistency of confirmed IRF-assessed ORR results when comparing the Atezo+Cabo arm to the Cabo arm will be investigated across subgroups, as defined in Section 5.3.4.1. The odds ratio of Atezo+Cabo versus Cabo and its 95% CI will be presented for each subgroup in a Forest plot.

5.8 INTERIM ANALYSES

5.8.1 Planned Interim Analyses

There are no planned interim analyses for the primary endpoint of IRF-PFS.

A total of three analyses of OS will be performed, including two interim analyses and one final analysis. The boundary for statistical significance at each OS analysis will be determined based on the Lan-DeMets implementation of the O'Brien-Fleming (OBF) function (DeMets and Lan 1994) to maintain the overall type I error rate (Hung et al. 2007; Glimm et al. 2010) at either 0.03 or 0.05 level, depending on whether primary endpoint of IRF-PFS is significant at 0.02 level. The OBF boundary for statistical significance is provided in Table 3. The OS endpoint will be considered positive in the ITT population if statistical significance is achieved in favor of the Atezo+Cabo arm for any of the two OS interim analyses or the final analysis.

Table 3 Analysis Timing and Stopping Boundaries for Overall Survival Interim and Final Analyses

	OS Interim Analysis 1	OS Interim Analysis 2	Final OS
Percent Information[a]	53% 175 events	80% 260 events	100% 325 events
Timing from FPI	27 months[b]	39 months	52 months
OBF Boundary when $\alpha = 0.03$ ($\alpha = 0.05$)	0.0019 (0.0045)	0.0125 (0.0231)	0.0259 (0.0424)
MDD when $\alpha = 0.03$ ($\alpha = 0.05$)	HR $\leq$ 0.62 (HR $\leq$ 0.65)	HR $\leq$ 0.73 (HR $\leq$ 0.75)	HR $\leq$ 0.78 (HR $\leq$ 0.80)

FPI = first patient in; HR = hazard ratio; MDD = minimum detectable difference; OBF = O'Brien-Fleming; OS = overall survival; PFS = progression-free survival.

[a] Corresponds to the number of death events required for the final analysis of OS

[b] At the time of PFS primary analysis

The first interim analysis of OS will be performed at the time of the PFS primary analysis. A total of 175 OS events are expected at the first interim analysis of OS, which corresponds to 53% of the events information required for the final analysis of OS in the ITT population. Statistical significance will be declared if $p < 0.0019$. If there are significantly fewer (< 160) OS events than the expected 175 OS events, then the first interim analysis will be delayed until 175 OS events occur. An administrative α of

0.000001 (negligible impact on overall type I error rate) will be spent on the OS hypothesis at the time of the planned PFS.

The second interim analysis of OS will be performed when approximately 260 deaths have occurred, which corresponds to approximately 80% of the events information required for the final analysis of OS in the ITT population. Statistical significance will be declared if $p < 0.0125$.

The final analysis of OS will be performed when 325 deaths (65% of 500 patients in the ITT population) have occurred. Statistical significance will be declared if $p < 0.0259$ when exactly 325 deaths have occurred at the time of the final OS analysis.

The actual OBF boundary will be calculated at the time of analysis based on actual number of events observed.

6. SUPPORTING DOCUMENTATION

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

7. REFERENCES

Bretz F, Maurer W, Brannath W, et al. A graphical approach to sequentially rejective multiple test procedures. Stat Med 2009;28:586–604.

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

Burman CF, Sonesson C, Guilbaud O, et al. A recycling framework for the construction of Bonferroni-based multiple tests. Stat Med 2009;28:739-61.

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

Glimm E, Maurer W, Bretz F. Hierarchical testing of multiple endpoints in group-sequential trials. Stat Med 2010;29:291-28.

Hung HM, Wang SJ, O'Neil R. Statistical considerations for testing multiple endpoints in group sequential or adaptive clinical trials. J Biopharm Stat 2007;17:1201-10.

Signature Page for Statistical Analysis Plan v2 - WO41994 - Published
System identifier: RIM-CLIN-450003

Approval Task	[redacted] Company Signatory 02-Sep-2022 14:28:36 GMT+0000