X184-312: Statistical Analysis Plan

A Randomized, Controlled Phase 3 Study of Cabozantinib (XL184) in Combination with Atezolizumab versus Sorafenib in Subjects with Advanced Hepatocellular Carcinoma Who Have Not Received Previous Systemic Anticancer Therapy
Version 1.0

Date: February 5, 2021

Prepared By	PPD Exelixis
Approved By	PPD Exelixis

Approved By	PPD  Exelixis
Approved By	PPD  PPD , Exelixis
Approved By	PPD  PPD , Exelixis

TABLE OF CONTENTS

LIST OF APPENDICES

LIST OF ABBREVIATIONS

AE	Adverse event
AESI	Adverse event of special interest
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
ATA	Adequate tumor assessment
ATC	Anatomical Therapeutic Chemical
BIRC	Blinded independent review committee
BMI	Body mass index
BUN	Blood urea nitrogen
CI	Confidence interval
CRF	Case report form
CT	Computerized tomography
CTC	Circulating tumor cell
CTCAE	Common terminology criteria for adverse events
CTMS	Clinical trial management system
DBP	Diastolic blood pressure
EBRT	External beam radiation therapy
ECOG PS	Eastern Cooperative Oncology Group Performance Status
ER	Emergency room visit
FDA	Food and Drug Administration
GGT	Gamma-glutamyltransferase
HCC	Hepatocellular carcinoma
HCRU	Health care resource utilization
HGB	Hemoglobin
HR	Hazard ratio
HRQOL	Health Related Quality of Life
ICH	International Conference on Harmonization
ICI	Immune checkpoint inhibitors
ICU	Intensive care unit
IDMC	Independent Data Monitoring Committee
IMM	Immune modulating medication
INR	Prothrombin International normalized ratio

ITT	Intent-To-Treat
IRT	Interactive Response Technology
LDH	Lactate dehydrogenase
LLQ	Lower limit of quantitation
MRI	Magnetic resonance imaging
MedDRA	Medical Dictionary for Regulatory Activities
NPACT	Non-protocol anticancer therapy
OS	Overall survival
ORR	Objective response rate
PD	Progressive Disease
PFS	Progression-free survival
PK	Pharmacokinetic
qd	Once daily
QOL	Quality of Life
SAE	Serious adverse event
SBP	Systolic blood pressure
SAP	Statistical analysis plan
SD	Stable Disease
TEAE	Treatment emergent-adverse event
TSH	Thyroid-stimulating hormone
UPCR	Urine protein/creatinine ratio
VAS	Visual analogue scale
WHO-DD	World Health Organization drug dictionary

1 ADMINISTRATIVE STRUCTURE AND VERSION HISTORY

Exelixis, Inc. is sponsoring this study. Exelixis is responsible for the statistical design and planning of this study. Statistical programming and analyses are being conducted under contract by Array Inc. in conjunction with Exelixis, Inc.

This version of the Statistical Analysis Plan (SAP) is based on the protocol amendment 2.0 dated April 09, 2020. The scope of this SAP is limited to the analyses of data pertaining to subjects enrolled in the global population (as defined in Section 2.1).

Table 1: Protocol Version History

Date	Version	Primary Reason(s) for Amendment
31 August 2018	Original Protocol	Not Applicable
12 April 2019	Amendment 1	Regulatory agency feedback for more robust characterization of the single-agent activity of cabozantinib.
09 April 2020	Amendment 2	Addition of Mainland China extension phase and an interim analysis for secondary PFS
23 June 2020	Addendum to Amendment 2	Increased enrollment to up to 840 to allow at least 185 to be enrolled in the single-agent arm per protocol specified assumption

Table 2: SAP Version History

Date	Version	Primary Reason(s) for Amendment
05 February 2021	1	Not Applicable

2 STUDY DESCRIPTION

2.1 Study Design

This is a multicenter, randomized, open-label controlled Phase 3 study of cabozantinib in combination with atezolizumab versus sorafenib in subjects with advanced hepatocellular carcinoma (HCC) who have not received previous systemic anticancer therapy. The multiple primary efficacy endpoints for the study are progression-free survival (PFS) assessed per blinded independent radiology committee (BIRC) and overall survival (OS) for the experimental arm (cabozantinib + atezolizumab) vs. the control arm (sorafenib). Additionally, there will be a third arm to evaluate the safety and activity of cabozantinib as a single agent. Approximately 840 eligible subjects with advanced HCC will be randomized in a 2:1:1 ratio at approximately

250 sites to receive cabozantinib plus atezolizumab combination, or sorafenib, or cabozantinib alone, respectively. After completion of the global enrollment phase, additional subjects (up to 148) may be enrolled in a Mainland China Extension Phase at sites in Mainland China for evaluation in a China subpopulation. The global population will include the first 837 subjects enrolled during the global enrollment phase (including subjects enrolled at sites in Mainland China during that phase), and the China subpopulation will include all subjects enrolled at sites in China (i.e., during both the global enrollment phase and the Mainland China Extension Phase).

Each subject's course of treatment will consist of the following study periods:

Pre-treatment Period: Potential subjects will be screened to determine if they meet the required eligibility criteria. Qualifying screening assessments must be performed within 28 days before randomization unless otherwise specified.

Treatment Period: Subjects who meet all study eligibility criteria will be randomly assigned in a 2:1:1 ratio to the following open-label study treatment arms, respectively:

Experimental arm (~370 subjects): cabozantinib 40 mg oral daily qd + atezolizumab 1200mg infusion every three weeks (q3w)
Control arm (~185 subjects): sorafenib 400mg oral twice daily (bid)
Single-Agent Cabozantinib arm (~185 subjects): cabozantinib 60 mg oral daily (qd)

Randomization will be stratified by the following factors established at screening:

- Disease etiology (HBV [with or without HCV], HCV [without HBV], or Other)
- Region (Asia, Other)
- Presence of extrahepatic disease and/or macrovascular invasion (Yes, No)

Subjects may receive study treatment (including after radiographic progression per RECIST 1.1) as long as they continue to experience clinical benefit in the opinion of the Investigator or until there is unacceptable toxicity, the need for subsequent systemic anticancer treatment, or any other reasons for the treatment discontinuation listed in the protocol. Subjects on the experimental arm (cabozantinib in combination with atezolizumab) may be allowed to discontinue one component of the study treatment but continue to receive the other for management of AEs if necessary. For subjects on the experimental arm, escalation of cabozantinib from 40 mg qd to 60 mg qd is allowed after Sponsor approval.

Crossover Phase: The study may transition to a Crossover Phase upon decision by the Sponsor and following discussion with regulatory authorities, if the analysis of OS (one of the multiple

primary endpoints) shows statistically significant and clinically meaningful evidence of improvement. The study is expected to have completed enrollment at the time of transitioning to the Crossover Phase. If the decision is made to enter the Crossover Phase, study sites will have 8 weeks to determine eligibility and begin administration of crossover treatment. Any crossover to experimental therapy will not be allowed unless study transitions to a Crossover Phase.

- Subjects randomized to the sorafenib control arm or the single-agent cabozantinib arm will have the option to cross over to receive the cabozantinib + atezolizumab combination if they meet predefined eligibility criteria
- Subjects randomized to the cabozantinib + atezolizumab experimental arm and subjects randomized to the sorafenib control arm or the single-agent cabozantinib arm who are still receiving study treatment and have not crossed over to the combination treatment (cabozantinib + atezolizumab) may continue on their originally assigned study treatment until a criterion for protocol-defined discontinuation has been met. During the post-treatment period they will continue with post treatment assessments.

Post-Treatment Period: Post-treatment follow-up visits FU-1 and FU-2 for safety evaluation will occur 30 (+14) days and 100 (±14) days, respectively, after the date of decision to permanently discontinue study treatment. In addition, subjects are to be contacted every 12 weeks (±14 days) after FU-2 visit to assess survival status and receipt of any subsequent anticancer therapy. See protocol 'Study Design' section for additional details.

Maintenance Phase: Upon determination by the Sponsor that sufficient data have been collected to adequately evaluate all study endpoints to establish the safety and efficacy profile of the study treatment for regulatory purposes, the study will commence transition to the Maintenance Phase.

If a Crossover Phase has been implemented, the Maintenance Phase may not begin before the Week 9 Day 1 (W9D1) visit has elapsed for the last subject who crossed over. Subjects will continue to receive study treatment until a criterion for protocol-defined discontinuation has been met. See protocol 'Study Design' section for details of maintenance phase assessments.

2.2 Study Treatment

Eligible subjects will be randomly assigned in a 2:1:1 ratio to the following open-label study treatment arms, respectively:

Experimental arm: cabozantinib 40 mg oral qd + atezolizumab 1200 mg infusion q3w

Single-agent cabozantinib arm: cabozantinib 60 mg oral qd
Control arm: sorafenib 400 mg oral bid (800 mg daily)

2.2.1 Study Objectives and Endpoints

The primary objective of this study is to evaluate the efficacy of cabozantinib in combination with atezolizumab versus sorafenib in subjects with advanced HCC who have not received previous systemic anticancer therapy. A secondary objective is to evaluate the single-agent activity of cabozantinib compared with sorafenib in this patient population.

This study has two primary efficacy endpoints. The trial will be declared a success if the null hypothesis is rejected for either of these endpoints; rejection of the null hypotheses for both endpoints is not required.

2.2.2 Primary Efficacy Endpoints

The two primary efficacy endpoints are:

- Duration of progression-free survival (PFS) per Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST 1.1) (Eisenhauer, 2009) by Blinded Independent Review Committee (BIRC) (see Section 7.4.1) for the experimental arm (cabozantinib + atezolizumab) vs. the control arm (sorafenib)
- Duration of overall survival (OS) for the experimental arm (cabozantinib + atezolizumab) vs. the control arm (sorafenib)

2.2.3 Secondary Efficacy Endpoints

- PFS per RECIST 1.1 by BIRC for the single-agent cabozantinib arm vs. the control arm (sorafenib)

2.2.4 Additional Endpoints

- ORR, TTP, and DOR per RECIST 1.1 by BIRC and Investigator
- Evaluation of radiographic response per modified RECIST (mRECIST)
- Safety through the evaluation of AEs, including immune-related adverse events (irAEs) and other AEs of special interest (AESIs).
- Characterization of the pharmacokinetics (PK) of cabozantinib in subjects with previously untreated HCC
- Immunogenicity of atezolizumab given in combination with cabozantinib
- Change in serum AFP from baseline
- Correlation of biomarker analyses with clinical outcomes
- Health-related quality of life (HRQOL) as assessed by the EuroQol Health questionnaire instruments (EQ-5D-5L)
- Health care resource utilization

2.3 Power and Sample Size Justification

The study is designed to provide adequate power for analyses of both primary endpoints of PFS and OS comparing the experimental arm (cabozantinib + atezolizumab) with the control arm (sorafenib), and for the analysis of the secondary endpoint of PFS comparing single-agent cabozantinib with the control arm. For the primary endpoints, a larger sample size is needed to provide reasonable power for OS than is required to evaluate PFS. As a result, if PFS were to be evaluated in the entire study sample, the PFS events may be biased toward shorter progression times. Thus, to allow longer, more robust PFS follow up among a smaller number of subjects, this study employs a "trial within a trial design" (Hessel C, 2016).

Per protocol amendment 2, the total sample size was 740 subjects, to be randomized in a 2:1:1 fashion: 370 to the experimental combination arm (cabozantinib + atezolizumab), 185 to the control arm (sorafenib), and 185 to the single-agent cabozantinib arm.

Per addendum to the protocol amendment 2, the sample size for the global study was increased to ensure the enrollment of at least 185 subjects in the single agent cabozantinib arm in order to allow for enrollment of the protocol-specified number of subjects in this arm. The analyses are to be carried upon observance of the specified number of events as stated in the protocol.

For PFS, a total of 257 events in the first 372 subjects randomized in a 2:1 ratio in the experimental and control arms (n=248 and 124 respectively; the PITT population) provide the study with 90% power for a 2-sided log-rank test with a 1% level of significance to detect a hypothesized true HR of 0.6. Assuming an exponential distribution of PFS, this corresponds with a 67% increase in median PFS from 3.6 months to 6.0 months. The minimum observed effect that would result in statistical significance for PFS is an HR of 0.71, a 41% improvement in median from 3.6 to 5.1 months. Interim analysis of PFS is not planned.

For OS, a total of 368 deaths among all 555 subjects randomized in a 2:1 ratio in the experimental (n=370) and control arms (n=185) are required to provide 90% power to detect an HR of 0.69 using the log-rank test and a 2-sided significance level of 4%. Assuming an exponential distribution for OS, this corresponds to a 45% increase in median survival from 12.3 months to 17.8 months. The minimum observed effect that would result in statistical significance for the primary analysis of OS is an HR of 0.7942, a 26% improvement in median from 12.3 to 15.5 months.

For the secondary endpoint of PFS, a total of 283 events among all 370 subjects (185 per arm) in the single-agent cabozantinib and control arms provides the study with 85% power for a 2-sided log-rank test with a 1% level of significance to detect a hypothesized true HR of 0.65. Assuming an exponential distribution for PFS, this corresponds with a 53% increase in median PFS from 3.6 months to 5.5 months. The minimum observed effect that would result in statistical significance for PFS is an HR of 0.736, a 36% improvement in median from 3.6 to 4.9 months.

With an average accrual rate of 30 subjects per month it is estimated that it will take approximately 18.4 months to observe the required primary PFS events (16.5 months of subject accrual) and approximately 38 months to observe the required deaths for OS (25 months of subject accrual) to evaluate the primary endpoints in the combination and control arms. Assuming the same accrual rate, it is anticipated that it will take approximately 25 months to observe the required secondary PFS events (25 months of subject accrual) to evaluate the secondary endpoint in the single-agent cabozantinib and control arms. These estimates are based upon the statistical assumptions, adjusted to accommodate the time required for all study sites to be activated and for subject recruitment to reach full potential. The true intervals required to meet these milestones may be longer or shorter due to divergence from assumptions, including non-constant accrual rate due to the time required for all study sites to become active. The sample size was increased to up to 837 subjects to ensure the needed enrollment of at least 185 subjects in the single-agent cabozantinib arm.

An overview of the endpoints and operating characteristics is shown in Table 3:

Table 3: Summary of Endpoint Operating Characteristics

Accrual per month	30 subjects		
Randomization allocation	2:1 within Experimental and Control Arms		1:1 within Single-Agent Cabozantinib and Control Arms
Endpoint:	**PFS: Primary endpoint**	**OS: Primary endpoint**	**PFS: Secondary endpoint**
Power	90%	90%	85%
Alpha allocated (2-sided)	0.01	0.04	0.01
# of interim analyses (approximate information fraction)	0 (NA)	2 (33%, 66%)	1 (67%)
Assumed median, control (months)	3.6	12.3	3.6
Assumed median, experimental (months)	6.0	17.8	5.5
Assumed HR	0.60	0.69	0.65
Number of events	257	368	283
N for analysis Population Planned n1 vs n2	372 PITT Population 248 vs 124	555 ITT Population 370 vs 185	370 ITT Population 185 vs 185
Time to enroll (months)	16.5	25	25
Time to trigger event (months)	18.4	38	25

Power and sample size estimates were calculated using EAST v6.5 by Cytel Software.

2.4 Randomization and Blinding

This is a multicenter, randomized, open-label controlled Phase 3 study. The study treatment assignment will be known to the subjects, investigators, study centers, Sponsor, and any Contract Research Organization affiliated with the study other than the BIRC. Study staff supporting the BIRC will ensure that the blind is maintained for the BIRC.

After obtaining informed consent, the site representative will use the designated web-based interactive response technology (IRT) system to register a subject for screening. The IRT will assign a unique subject number. When a subject has been deemed eligible at the study site, the site representative will use the IRT to randomize the subject into the study.

Eligible subjects will be randomly assigned in a 2:1:1 ratio to the three study treatment arms. (See Section 2.2):

Stratified randomization will be conducted using permuted blocks over 12 strata. The 12 strata are based on all combinations of the following 3 stratification factors which will be established at screening:

- Disease etiology (HBV [with or without hepatitis C virus {HCV}], HCV [without HBV], or Other)
- Region (Asia, Other)
- Presence of extrahepatic disease and/or macrovascular invasion (Yes, No)

Protocol amendment 1 necessitated the sample size of the single-agent cabozantinib arm to be increased to support the comparison of the single-agent cabozantinib arm with the sorafenib (control) arm. As a consequence, the original allocation ratio of 6:3:1 of the experimental arm, control arm and the single-agent cabozantinib arm was changed to 2:1:1 in the respective arms in the amendment. The subjects enrolled under the original protocol will have the initial allocation ratio of 6:3:1.

3 ANALYSIS POPULATIONS

The planned analysis populations are briefly summarized in Table 4. The populations specified below pertain to the global study population and do not include the subjects enrolled in the China Extension Phase. The addendum to Protocol amendment 2 allowed for the global enrollment (excluding the China Extension Phase) to be up to 840 subjects to ensure enrollment of at least 185 subjects in the single agent cabozantinib arm. The study enrolled 188 subjects in the single agent cabozantinib arm in the global study population. The 188th subject randomized to the single agent cabozantinib arm was considered the last subject in the global study population of 837 subjects. Subjects randomized in China after this date are in the China Extension Phase only. Analyses that include the China Extension Phase are beyond the scope of this plan. Further information on the global study population is provided in the sections below.

Table 4: Analysis Populations for the Global Study Population

Analysis	Population	Subjects
Primary PFS	PITT	First 372 subjects randomized to experimental or control arms
Primary OS	ITT	All subjects randomized to experimental or control arms
Secondary PFS	ITT	All subjects randomized in control and single-agent cabozantinib arms
Additional efficacy analyses	ITT	All randomized subjects in the three treatment arms
Safety analyses	Safety	All subjects who received any study treatment
Safety analyses for PITT subjects	P-Safety	All subjects who received any study treatment in the PITT population

All populations exclude the China Extension Phase.

ITT, Intent-to-Treat; OS, overall survival; PFS, progression-free survival; PITT, PFS Intent-to-Treat; P-Safety, PFS Safety

3.1 Intent to Treat Population

The ITT population will consist of all subjects who are randomized to any of the three study treatment arms (the experimental arm [cabozantinib + atezolizumab], the single-agent cabozantinib arm, or the control arm [sorafenib]), regardless of whether any study treatment or the correct study treatment was received.

3.2 PFS Intent to Treat Population

The first 372 subjects randomized (based upon Greenwich Mean Time randomization date/time values) to the experimental (cabozantinib + atezolizumab) arm and control (sorafenib) arm will be considered as the PFS Intent-to Treat (PITT) population. This population will be used for the analysis of the primary endpoint. The population may be extended to the first 410 subjects randomized to the two groups if a review of the accumulating events required will not be reached (due to censoring) among the first 372 subjects originally planned.

3.3 Safety Population

The Safety population will consist of all subjects who receive any amount of study treatment. Analyses based on Safety population will be performed according to the actual treatment received for the length of the study. For example, a subject randomized to the sorafenib arm who received cabozantinib + atezolizumab in error, and only received the combination regimen, would be analyzed in the combination arm for safety summaries. Subjects who receive treatment components from both the cabozantinib + atezolizumab and sorafenib arm in error (e.g. received sorafenib and either cabozantinib and/or atezolizumab) will be excluded from the safety population. Their safety data will be presented in by-subject listings.

3.4 PFS Safety Population

The PFS Safety (P-Safety) population will include the subjects included in the PITT population receiving any amount of study treatment.

3.5 Per Protocol Population

A Per Protocol population is not defined or planned for the study.

4 GENERAL CONVENTIONS

The statistical principles applied in the design and planned analyses of this study are consistent with International Conference on Harmonization (ICH) E9 (ICH Harmonized Guideline E9 Statistical Principles for Clinical Trials, 1998) and draft E9 R1 guidelines (ICH Harmonized Guideline E9 (R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials (Draft), 2017).

Continuous data will be summarized using descriptive statistics (number of observations, mean, standard deviation, median, 25^{th} and 75^{th} percentiles, minimum, and maximum). Frequencies and percentages will be used for summarizing categorical (discrete) data.

Confidence intervals, when presented, will generally be constructed at the 95% level. For binomial variables, the Clopper-Pearson method (Clopper CJ, 1934) may be employed unless otherwise specified.

A month is operationally defined to be 30.4375 days. Six months is operationally defined to be 183 days.

Efficacy summaries will be presented by treatment arm randomized to and safety summaries by treatment received unless otherwise specified.

4.1 Treatment Periods, Summary Groups, Analyses, and Reports

For each subject, data collected for this study will be grouped by the following two treatment periods in which it was obtained:

Core Period: This period begins with screening. The end of this period is the earlier of the data cutoff date (for a particular analysis), date of withdrawal from study, date of discontinuation from or loss to survival follow-up, date of death, date of initiation of the Maintenance Phase or date of enrollment to the Crossover Phase -1.

Crossover Period: The start of this period is the enrollment date of the Crossover Phase, and it ends at the earlier of date of withdrawal from study, date of discontinuation from study treatment or loss to survival follow-up, date of death, or date the Maintenance Phase is initiated.

The following sets of summaries are planned:

- Primary safety and efficacy summaries for the clinical study report (CSR) and any CSR addendums in support of primary efficacy endpoint analyses will include all and only data obtained in the Core Period and presented separately as follows:
  - cabozantinib + atezolizumab versus sorafenib
  - single-agent cabozantinib versus sorafenib

Table 5: Schema for Efficacy Summaries

Treatment Period	**Core Period (Summaries provided in primary CSR and efficacy addendum(s))**		
Population	Subjects randomized to cabo+atezo (Experimental Arm)	Subjects randomized to single-agent cabozantinib (Single-agent arm)	Subjects randomized to sorafenib (Control arm)
Start Date of Period	Randomization date	Randomization date	Randomization date

- Upon completion of all study data collection, only safety summaries will be repeated for the final CSR safety addendum and will include data from the Crossover period and updated data from the Core period

Table 6: Schema for Safety Summaries

Treatment Period	Core Period (Summaries provided in primary CSR and updated in CSR safety addendums)			Crossover Period (Summaries provided in final CSR safety addendum only)
Population	Subjects randomized to and who received cabozantinib+atezolizumab (Experimental Arm)	Subjects randomized to and who received single-agent cabozantinib (Single-agent Arm)	Subjects randomized to and who received sorafenib (Control arm)	sorafenib or single agent cabozantinib subjects who cross over to receive cabozantinib+atezolizumab (Crossover subjects)
Start Date of Period	Date of 1st dose of cabo and/or atezo	Date of 1st dose of cabozantinib	Date of 1st dose of sorafenib	Date of 1st dose of cabo+atezo

Up to six study analyses are planned over the course of the study:

1. Primary PFS + interim OS #1 + interim secondary PFS [Core treatment period]
2. Interim OS #2, if interim OS #1 is not successful [Core treatment period]
3. Final OS, if interim OS #1 and #2 are not successful [Core treatment period]
4. Final secondary PFS analysis [Core treatment period]
5. Safety update for 90-day report [Core treatment period]
6. Final safety update, upon initiation of the Maintenance Phase [Core and Crossover treatment periods]

Some of the analyses listed above may be contemporaneous depending on the occurrence of the events. Details on the interim analyses of OS, secondary PFS and the control of type 1 error are provided in Sections 7.2 and 7.3.

4.2 Definition of Baseline

In general, for efficacy endpoints the last observed measurement prior to randomization will be considered the baseline measurement.

For safety endpoints the last observation before first dose of study treatment will be considered the baseline measurement unless otherwise specified and will be established for each period defined in Section 4.6.

For assessments on the day of first dose of study treatment where time is not captured, if such procedures are required by the protocol to be conducted before first dose or a nominal pre-dose indicator is available, this will serve as sufficient evidence that the assessment occurred prior to first dose.

For the experimental arm (cabozantinib + atezolizumab treatment), the earliest instance of administration of either cabozantinib or atezolizumab on the study will be considered as the time point of first dose. Similarly, the latest instance of administration of either of the treatments will be considered as the time point of last dose.

4.3 Definition of Study Day

For the purpose of efficacy data summaries which are based only on data collected in the Core period, Study Day is defined with respect to the randomization date. For visits (or events) that occur on or after randomization, study day is defined as (date of visit [event] – date of randomization + 1). For visits (or events) that occur prior to randomization, study day is defined as (date of visit [event] – date of randomization). There is no Study Day 0.

For the purpose of safety data summaries, Dose Day is defined with respect to the date of first dose of study treatment received in the respective periods defined in Section 4.6. For visits (or events) that occur on or after first dose, dose day is defined as (date of visit [event] – date of first dose of study treatment + 1). For visits (or events) that occur prior to first dose, dose day is defined as (date of visit [event] – date of first dose of study treatment). There is no Dose Day 0.

The last dose date for each individual agent and for any study treatment is defined for each of the periods described in Section 4.1. For listings (such as for adverse events [AEs]) that include the derivations of "days since last dose," this is defined for each agent as (event date – date of last dose for the agent). Two such fields, one for cabozantinib and one for atezolizumab will be presented for subjects receiving the combination of these agents. Events that occur on the same day as the last dose of a particular study drug will therefore be described as occurring zero days from the last dose of that study drug.

4.4 Visit Window Calculation

Analyses will be according to actual visit dates and times. The planned analyses do not require the calculation of visit windows. However, for analyses that require a particular visit as planned, measurements included in that visit must have occurred during the acceptable window defined for the visit in the protocol, unless otherwise specified in this plan.

4.5 Missing and Partial Data

In general, other than for partial dates, missing data will not be imputed and will be treated as missing. The algorithms for imputation of partial dates vary depending upon the parameter. These are presented in Appendix A.

4.6 Safety Observation Periods

Generally only the safety data (including adverse events that start or worsen, laboratory results, vital signs, ECGs, ECOG PS, concomitant medications, etc.) reported during a safety observation period (defined below) will be analyzed and summarized, unless otherwise specified in this plan.

4.6.1 Standard Observation Period

Two standard safety observation periods are defined: one for the Core period and one for Crossover period.

Core Treatment Period: the standard safety observation period is the interval from the first dose date of original study treatment to the earliest of:

- Date of last dose of last component of original study treatment + 30 days or
- Date of death or
- Date subject withdrew consent for non-interventional study assessments or
- Data cutoff date (for a particular analysis) or
  Date of enrollment in the Crossover Phase - 1

Crossover Treatment Period: the standard safety observation period is the interval from the first dose date of crossover study treatment to the earliest of:

- Date of last dose of last component of crossover study treatment + 30 days or
- Date of death or
- Date subject withdrew consent for non-interventional study assessments or
- Data cutoff date (for a particular analysis)

4.6.2 Extended Observation Period

To evaluate the incidence of adverse events of special interest (AESIs), extended observation periods are defined for summarizing safety data as below.

Core Treatment Period:

- Date of last dose of last component study treatment + 100 days or
- Date of death or
- Date subject withdrew consent for non-interventional study assessments or
- Data cutoff date for analysis under consideration

Crossover Treatment Period:

- Date of last dose of last component study treatment + 100 days or
- Date of death or
- Date subject withdrew consent for non-interventional study assessments or
- Data cutoff date for analysis under consideration or
- Date of initiation of maintenance phase -1

4.7 Definition of Prior, Concomitant, and Subsequent Therapy

For the purpose of inclusion in summary tables, incomplete medication or radiation start and stop dates will be imputed as detailed in Appendix A. Based on imputed start and stop dates:

- Prior medications/radiation therapies are defined as medications with a stop date occurring before the date of first dose of study treatment.

- Subsequent medications/radiation therapies are defined as those that start after the date of randomization.

- Concomitant medications/radiation therapies are defined as medications that stop or continue on or after date of first dose day through the end of safety observation period.

Medications/radiation therapies may be summarized as prior, concomitant and/or subsequent.

4.8 Software

All analyses will be conducted using SAS Version 9.3 or higher.

4.9 Changes to Planned Analyses

Substantive changes to the analyses described in the protocol or in approved versions of this plan will be fully documented in a revised version of this plan approved by the Sponsor prior to conducting analyses that may be potentially impacted by said changes.

Clarifications, minor corrections, and operational considerations necessary to accurately conduct the analyses that do not materially change the nature of the analysis will be documented in an addendum to this plan that will be also be approved by the Sponsor prior to final study analyses.

Minor typographical error in the protocol are corrected in Table 11 of this plan.

5 STUDY POPULATION SUMMARIES

5.1 Enrollment

Subjects are defined to be enrolled at randomization. Enrollment will be summarized by region, country, site, and protocol version for the PITT and ITT populations.

5.2 Disposition

Subject disposition will be summarized categorically and will include the number and percentage of subjects in the ITT, PITT, P-Safety and Safety populations.

The reasons for each study treatment discontinuation, discontinuation of radiographic follow-up and study follow-up discontinuation will also be summarized categorically. Additional summaries of the combination arm will also be provided to characterize the status and nature of subjects discontinuing treatment components due to AE.

All screen failure subjects will be summarized with the reason for screen failure.

5.3 Demographic and Baseline Characteristics

Summaries of demographics, stratification factors and baseline characteristics will be presented for subjects in the ITT, PITT, P-Safety and Safety populations.

[A] The demographic characteristics include:

- Age (continuous)
- Age category 1
  - < 65 years
  - ≥ 65 years
- Age category 2
  - < 75 years
  - ≥ 75 years
- Age category 3
  - <65 years
  - 65 to <75 years
  - 75 to <85 years
  - ≥85 years

- Sex
  - Male
  - Female
  - Not reported
- Ethnicity
  - Hispanic or Latino
  - Not Hispanic or Latino
  - Not Reported
- Race
  - American Indian or Alaska Native
  - Asian
  - Black or African American
  - Native Hawaiian or Other Pacific Islander
  - White
  - Not Reported
  - Other
- Geographic Region
  - Asia
  - Other

Note for this study birth date is not collected but age in years is collected at informed consent.

[B] Categorical summaries of the following stratification factors will be presented as recorded (a) in the IRT during randomization (b) on the CRF (c) cross tabulation of all 3 stratification factors per IRT (d) cross tabulation of all 3 stratification factors per CRF (e) cross-tabulation of geographic region and etiology of disease per CRF:

- etiology of disease
  - HBV [with or without HCV]
  - HCV [without HBV]
  - Other
- geographic region (Asia, Other Regions)
- presence of extrahepatic spread of disease and/or macrovascular invasion (Yes, No)

[C] Baseline characteristics include:

- Height in cm– descriptive statistics

- Weight in kg – descriptive statistics
- Body mass index (BMI) in kg/meter^2, calculated as (weight in kg*10000)/(Height in cm)^2 –descriptive summary:
- ECOG PS: 0, 1, Missing
- Smoking history
    - Current
    - Former
    - Never
- Alcohol use – Categorical summary for subjects classified as current user, former user or never will be presented

5.4 Medical History

General medical history data will be coded per MedDRA.

5.5 Cancer History and Current Disease Status

Cancer history and current disease characteristics data collected on the cancer history CRF will be summarized categorically or with descriptive statistics as appropriate for the PITT and ITT populations. The following summaries are planned:

- Diagnosis of carcinoma of HCC by histology or cytology (Yes, No)
- Diagnosis of carcinoma of HCC by imaging (Yes, No)
- Current etiology:
    - Hepatitis B virus (without HCV)
    - Hepatitis C virus (without HBV)
    - Hepatitis B and C virus
    - Hepatitis B virus (regardless of HCV)
    - Hepatitis C virus (regardless of HBV)
    - Alcoholism
    - Cirrhosis
    - Nonalcoholic Steatohepatitis (NASH)
    - Non-alcoholic fatty liver disease (NAFLD)
    - Other

- Hepatitis B virus surface antigen (HBsAg), Hepatitis B virus core antibody (HBCAB) and HCV laboratory results summary
    - HBsAg (+) , HBCAB(+)
    - HBsAg (+) , HBCAB(-)
    - HBsAg (-) , HBCAB(+)
    - HBsAg (-) , HBCAB(-)
    - HCV (+)
    - HCV (-)
- BCLC Grade (See XL184-312 Protocol Appendix J)
    - 0 (Very Early Stage)
    - A (Early Stage)
    - B (Intermediate Stage)
    - C (Advanced Stage)
    - D (Terminal Stage)
- Child-Pugh Grade (See XL184-312 Protocol Appendix K)
    - A (score 5 - 6)
    - B (score 7 – 9)
    - C (score 10 – 15)
- Hepatic Encephalopathy Status for Child-Pugh Grade
    - None
    - Grade I-II
    - Grade III-IV
    - Missing or unknown
- Ascites Status for Child-Pugh Grade
    - Absent
    - Slight
    - Moderate
    - Tense
    - Missing or unknown
- Time in years to randomization since initial diagnosis of HCC (Note: Incomplete diagnosis dates will be imputed as detailed in Appendix A)
- Currently has metastatic disease per investigator (Yes, No)
- Current Extent of HCC Disease per CRF:
    - Portal Vein Invasion (Yes, No, Unknown)
    - Bile Duct Invasion (Yes, No, Unknown)
    - Macrovascular Invasion (Yes, No, Unknown)

    - Extrahepatic Disease (Yes, No, Unknown)
    - Other (Yes, No)

- Summaries for tumor assessment at screening per Investigator and BIRC are:
    - Extent of disease per target/non-target lesions
        - Liver
        - Bone
        - Lymph nodes
        - Important visceral sites
            - Adrenal gland
            - Lungs
        - All other sites
            - Incidence of all other sites will be individually reported
    - Number of target lesions (1, 2, ≥ 3)
    - Descriptive statistics for sum of target lesion diameter
    - Has measurable disease (Yes, No)
    - Number of organs with lesions at screening (1, 2, ≥ 3)
- PD-L1 immunohistochemistry status (High, Low/Negative, Unknown)

5.6 Baseline Laboratory Values

Frequencies and percentages will be provided for baseline laboratory values for selected laboratory parameters specified below for the PITT and ITT populations.

- Alpha-fetoprotein (AFP) (ng/mL)
    - <400
    - ≥400
- Prothrombin INR
    - ≤2.3
    - >2.3
- Serum albumin (g/L)
    - <35
    - ≥35

- Total bilirubin (μmol/L)
    - <22.23
    - ≥22.23 - <29.07
    - ≥29.07
- Platelets (10^9/L)
    - <80
    - ≥80
- ANC/lymphocyte ratio
    - <3
    - ≥3
- ALBI grade

ALBI grade will be defined based on the calculated ALBI score. ALBI Grade =1 if ALBI score ≤ -2.60, ALBI Grade=2 if ALBI score > -2.60 to ≤ -1.39, ALBI Grade=3 if ALBI score > -1.39.

ALBI Score = (log10 bilirubin[μmol/L] x 0.66) + (albumin[g/L] x (-0.0852))

- 1
- 2
- 3

6 TREATMENTS AND MEDICATIONS

6.1 Prior Non-Radiation Anticancer Therapy

The study aims to enroll subjects who have not received previous systemic anticancer therapy for advanced HCC including chemotherapy, small molecule kinase inhibitors, and immune checkpoint inhibitors (ICI). Subjects who have received local intra-tumoral or arterial chemotherapy are eligible.

The following will be summarized categorically or with descriptive statistics as appropriate for all subjects in the ITT, PITT, Safety and P-Safety populations:

- The time from the end of most-recent local non-radiation anticancer regimen for advanced HCC to randomization will be summarized descriptively
- Number of prior local non-radiation anticancer regimens (e.g. TACEs) for advanced HCC per subject (1, 2, ≥3) and descriptive statistics

All prior local non-radiation anticancer agents will be summarized categorically by Anatomical Therapeutic Chemical (ATC) Class Text and WHO-DD base substance preferred name by study treatment arm and context of therapy (HCC - Local, HCC – Neoadjuvant, HCC – Adjuvant, and Other). All prior systemic non-radiation anticancer agents will also be summarized, primarily to demonstrate that there were no subjects in the study who received such therapy.

6.2 Prior Radiation Therapy

Data obtained from the history of radiation therapy CRF will be summarized categorically or with descriptive statistics as appropriate for all subjects in the ITT, PITT, and Safety populations:

- Number of prior radiation therapies for HCC per subject (1, 2, ≥3) and descriptive statistics
- Subject incidence of radiation therapy by indication (HCC and Other)
- Subject incidence of radiation therapy by type (External beam radiation therapy [EBRT], Internal radiation therapy [Brachytherapy], Radioisotope therapy, Radioembolization, and Other) for HCC or Other indications
- Subject incidence of radiation therapy by site (Bone, Soft-tissue, Unknown) for HCC or Other indications
- Number and percent of subjects that received radiation by anatomic sites

6.3 Prior and Concomitant Medications (Excluding Anticancer Therapy)

Medications recorded on the CRFs will be coded using the WHO-DD. Prior and concomitant medications, other than prior and subsequent anticancer therapies, will be summarized by treatment group in the P-Safety and Safety populations by ATC and WHO-DD base substance preferred name. In addition, prior medications will also be summarized in the PITT and ITT populations by ATC and WHO-DD base substance preferred name. A listing of anti-viral therapies will be provided for the Safety population. Anticancer therapies are addressed in Sections 6.8and 6.9 of this plan.

6.4 Subsequent Immune Modulating Medications

Immune modulating medications (IMMs) are selected medications with immune-modulating effects that may be used to manage immune-related AEs that result from treatment with atezolizumab. A list of the verbatim terms for these medications will be identified and provided along with the ATC and WHO-DD base substance preferred name. The incidence of subsequent

IMM and high-dose IMM (defined as ≥ 40 mg/day of prednisone or equivalent) for AESIs will be summarized by treatment group in the Safety and P-Safety populations by ATC and WHO-DD base substance preferred name.

6.5 Study Treatment Exposure

Study treatment exposure will be summarized descriptively and will include subjects in the P-Safety and Safety populations. Due to different treatment types (infusion vs. oral tablet) and frequency of administration (q3wk vs daily) between atezolizumab and cabozantinib, the exposure summaries for cabozantinib and atezolizumab will be presented separately along with summary for study treatment exposure. The study protocol allows subjects on the combination treatment to discontinue one component of the combination treatment but continue to receive the other component. Date of discontinuation along with reason for discontinuation will be recorded on separate CRFs for cabozantinib and atezolizumab treatment components. Duration of cabozantinib is defined as time from first dose to date of decision to discontinue and duration of atezolizumab is defined as time from first dose to date of last dose. Inclusion of the various treatment arms in the presentation of exposure summaries are described in Section 4.6.

For the combination regimen (experimental arm) an overall summary will be provided as follows:

- Duration (in months) of exposure, defined as maximum of duration of cabozantinib exposure and duration of atezolizumab exposure, calculated as maximum of [(date of decision to discontinue cabozantinib/cut-off date – date of first dose of cabozantinib + 1) /30.4375 and (last dose date of atezolizumab – date of first dose of atezolizumab + 1) /30.4375]
- The number and percent of subjects for who received at least one non-zero dose of cabozantinib > 4 weeks after last dose of atezolizumab.
- The number and percent of subjects for who received at least one dose of atezolizumab > 4 weeks after non-zero dose of cabozantinib.

The other summaries for cabozantinib and atezolizumab given in combination will present each agent separately as the former is an oral drug and the latter is an infusion drug.

The following independent summaries will be provided for cabozantinib 40 mg (as part of combination), cabozantinib 60 mg (as a single agent) and sorafenib. The start and end dates refer to the respective individual agent.

- Duration (in months) of exposure, including dose holds, calculated as (date of decision to discontinue study treatment/cut-off date – date of first dose + 1) /30.4375
- Average daily dose (mg/day) of study treatment, calculated as (total mg received / duration of exposure in days)
- Percent dose intensity, calculated as 100*(average dose in mg/day) / (40 mg/day) for cabozantinib in experimental arm
- Percent dose intensity, calculated as 100*(average dose in mg/day) / (60 mg/day) for cabozantinib in single agent arm
- Percent dose intensity, calculated as 100*(average dose in mg/day) / (800 mg/day) for sorafenib (control arm).
- Duration (in months) of exposure, excluding dose holds, defined as (date of decision to discontinue study treatment – date of first dose – total days with dose received of 0 mg during this interval + 1) /30.4375
- The last dose level received (with or without dose holds)

Summaries for atezolizumab (experimental arm):

- Duration of exposure calculated as (last dose date infusion of atezolizumab – date of first dose of atezolizumab +1)/30.4375
- Number of infusions of atezolizumab received
- Number of expected infusions of atezolizumab defined as duration of exposure divided by 21
- Number of actual infusions / Number of expected infusions
- Average dose (mg/infusion) of atezolizumab treatment received, calculated as (total mg received / # of infusions of atezolizumab received.)
- Percent dose intensity calculated as 100*(total mg received)/((last dose date - first dose date + 21)*(1200/21))

6.6 Study Treatment Modifications

Treatment modifications (holds, delays, and reductions) for study treatment will be summarized in the P-Safety and Safety populations. Only modifications due to AE will be summarized by protocol-defined dose levels. Per protocol, cabozantinib and sorafenib doses can be reduced,

held, or discontinued, while atezolizumab infusions can be incomplete, interrupted, delayed, or discontinued but the dose cannot be reduced. Therefore, for the experimental arm, dose reduction summaries apply only to the cabozantinib treatment component. For summarization, atezolizumab delays will be considered dose holds and will also be counted as dose modifications. Incomplete or interrupted atezolizumab infusions will not be considered dose modifications and will be presented separately. Therefore, for the experimental arm (cabozantinib + atezolizumab), dose hold and dose modification summaries will be presented by individual treatment components as well as by combination.

The protocol allows only two levels of dose reduction for cabozantinib and sorafenib as shown below:

Table 7: Protocol Defined Dose Levels for Cabozantinib and Sorafenib

Study Arm Treatment	Dose Level Increase	Assigned Starting Dose	First Dose Level Reduction	Second Dose Level Reduction
Experimental arm cabozantinib	60 mg (qd)	40 mg (qd)	20 mg (qd)	20 mg (qod)
Single-agent arm cabozantinib	NA	60 mg (qd)	40 mg (qd)	20 mg (qd)
Control arm sorafenib	NA	800 mg (tdd)	400 mg (tdd)	400 mg (qod)

NA, not applicable; qd, once daily; qod, every other day; tdd, total daily dose.

A. The following summaries will be presented for Cabozantinib/Sorafenib:

i. For dose reductions due to AE

Categorical summaries for:

- Number of subjects with any dose reduction
- Number of subjects who received each dose level category
- Lowest non-zero dose level category received

Descriptive statistics for:

- Duration of treatment in months for each dose level category, defined for each subject as (cumulative number of days treatment was taken at each dose level) / 30.4375
- Time from first dose to first dose level reduction (days)
- Time from first dose to second dose level reduction (days)

ii. Summaries for dose holds due to AE:

- Descriptive statistics for number of dose holds (0 mg dose level) due to an AE per subject and classified as 1, 2, 3 and >3
- Descriptive statistics for total duration of dose holds per subject due to an AE and classified as ≥ 7 days, ≥ 14 days, and ≥ 21 days. Duration of each dose hold is calculated as (stop date of hold – start date of hold + 1)
- Frequency counts and percentages for subjects with duration of holds due to an AE classified as any number of days, ≥ 7 days, ≥ 14 days, and ≥ 21 days
- Descriptive statistics for duration of dose holds per dose hold and per subject due to an AE, calculated as (stop date of hold – start date of hold + 1)
- Descriptive statistics for time to first dose hold, time to first dose hold that was ≥ 7 days, ≥ 14 days, and ≥ 21 days. The time to dose hold is calculated as (start date of the hold – first dose date + 1)
- Descriptive statistics for time to second dose hold, time to second dose hold that was ≥ 7 days, ≥ 14 days, and ≥ 21 days

iii. Summaries for dose modifications (defined as a reduction or hold) due to AE:

- Frequency counts and percentages for subjects with any dose modifications
- Descriptive statistics for number of dose modifications (0-4)
- Descriptive statistics for time to the first dose modification
- Descriptive statistics for time to the second dose modification

The protocol allows (subject to Sponsor approval) escalation of cabozantinib dose from 40 mg qd to 60 mg qd for subjects in the experimental arm depending tolerance of the 40 mg dose level for at least 4 weeks. Summaries will be provided for number of subjects in the experimental arm who have their cabozantinib dose increased from 40 mg (qd) to 60 mg (qd). The duration on the 60 mg (qd) dose will also be summarized for these subjects. Note that any subsequent reduction from the sponsor-approved 60 mg (qd) dose level due to AE will be summarized separately as detailed earlier in this section with up to 3 levels of reductions.

B. The following summaries will be presented for Atezolizumab:

- Number of subjects with any incomplete infusion
- Number of subjects with any interrupted infusion
- Number of subjects with dose delay due to an AE (0, 1, 2, 3 and >3)
- Frequency counts and percentages for subjects with total duration of delay due to AE classified as any number of days, ≥ 7 days, ≥ 14 days, and ≥ 21 days
- Frequency counts and percentages for subjects with duration of delay due to an AE classified as any number of days, ≥ 7 days, ≥ 14 days, and ≥ 21 days
- Descriptive statistics for time to the first dose delay
- Descriptive statistics for time to the second dose delay

6.7 Study Treatment Non-Compliance and Dosing Errors

Treatment non-compliance and dosing errors for reasons other than AE will be summarized in the P-Safety and Safety populations. Frequency counts and percentages will be presented by treatment groups for:

- Subjects with dose hold (0 mg) due to non-compliance
- Subjects who received dose > maximum allowed dose level at any time (60 mg qd for cabozantinib, even for subjects who started at 40 mg qd; 800 mg for sorafenib)
- Subjects who received non-protocol specified dose level (≤ maximum allowed dose level) at any time due to non-compliance
- Subjects who received non-protocol specified dose level (≤ maximum allowed dose level) at any time due to site/logistic error or other reason

6.8 Non-Protocol Anticancer Therapy (NPACT)

For the purpose of supporting safety evaluations:

Subsequent non-radiation NPACT will be summarized by ATC text and WHO Drug base substance preferred name in the P-Safety and Safety populations.

For the purpose of supporting efficacy evaluations:

Subsequent NPACT, including radiation therapy, will be summarized by treatment group in the PITT and ITT populations as follows:

- Based on the non-radiation therapy received subjects will be categorized into one or more of the following categories: systemic, local, or unknown and all NPACTs falling under these categories will be summarized by ATC text and WHO Drug based substance preferred name
- Time to first systemic NPACT will be summarized by descriptive statistics
- Frequency counts and percentages will be presented for radiation therapy indication, type, and site

6.9 Concomitant and Subsequent Surgeries/Procedures

Post-randomization radiation therapy and surgery/procedures that impacted the target lesion(s) (Yes, No, Unknown) will be summarized by treatment group for subjects in the PITT, ITT, P-Safety and Safety populations.

6.10 Concomitant Transfusions

Concomitant transfusions will be summarized by transfusion type and treatment group for subjects in the P-Safety and Safety populations.

7 EFFICACY ANALYSES

Efficacy summaries for primary, secondary, and additional endpoints will include only data collected in the Core period. No efficacy analyses will be conducted on data collected in the Crossover Phase.

7.1 Primary Efficacy Endpoints

The two primary efficacy endpoints for this study are duration of PFS per BIRC and duration of OS. Formal hypothesis tests are planned for these endpoints.

7.1.1 Duration of Progression-Free Survival (PFS)

7.1.1.1 Definitions and Conventions

Only adequate tumor assessments (ATAs) will be considered in the determination of radiographic progression and censoring dates. For the purpose of this study, an ATA is defined as one that results in a time point assignment of complete response, partial response, stable disease/(non-CR, non-PD), or progression.

The recorded date of radiographic progression is the date of the tumor assessment visit at which progression is declared. If multiple scan dates are associated with a tumor assessment visit, the earliest scan date within the set will be chosen as the progression date.

7.1.1.2 Primary Estimand

The primary estimand for PFS is the difference in survival functions between treatment conditions in the duration of radiographic progression-free survival in the targeted population:

- irrespective of whether the assigned study treatment was given
- irrespective of clinical deterioration
- irrespective of whether local radiation was given to bone
- irrespective of surgical resection of non-target lesions
- in absence of any surgery to resect tumor lesions
- in absence of any other systemic or local non-protocol anti-cancer treatment
- in absence of any local radiation to soft tissue
- in absence of missing pre-randomization tumor assessment
- in absence of two or more missing consecutive adequate tumor assessments

Derived as follows:

Estimand attribute [1]	**Primary definition**		**Rationale (as needed)**
Population	Subjects randomized into the study intended to include patients with advanced HCC who have not received previous systemic anticancer therapy.		The initial clause "Subjects randomized into the study intended to include…" is included to align the estimand with the application of the ITT principle expected by regulatory reviewers
Endpoint	Duration of radiographic progression-free survival		
Intercurrent events	**Event**	**Strategy**	**Rationale (as needed)**
	Receipt of non-assigned study treatment or non-receipt of study treatment	Treatment policy	This strategy applied to align the estimand with the application of the ITT principle expected by regulatory reviewers.
	Clinical deterioration	Treatment policy	This strategy applied due to the nature of the primary endpoint, which is radiographic PFS.
	Receipt of local radiation to bone	Treatment policy	This strategy applied as local radiation to bone is typically palliative and does not directly confound radiographic evaluations of soft tissue.
	Surgical resection of non-target tumor lesions	Treatment policy	This strategy applied as target lesions are the primary focus of RECIST 1.1 evaluations.
	Surgical resection of target tumor lesions	Hypothetical	This strategy arises as a consequence of the convention of censoring for these intercurrent events that confound radiographic tumor assessments per RECIST 1.1., and the assumption in the Kaplan-Meier model that censored subjects would have behaved in a fashion similar to those not censored, i.e. not experienced these intercurrent event(s)
	Receipt of systemic non-protocol anti-cancer medications	Hypothetical	
	Receipt of local non-protocol anti-cancer medications	Hypothetical	
	Receipt of local radiation to soft tissue for disease under study	Hypothetical	
	Missing pre-randomization tumor assessment	Hypothetical	
	Missed two or more consecutive adequate tumor assessments immediately prior to rPD or death	Hypothetical	
Population summary	Difference in survival functions between treatment conditions.		

[1]See Appendix E for estimand terminology

7.1.1.3 Alternative Estimands

Two alternative estimands for PFS are defined as below, arising from changes in strategy for handling some intercurrent events. Shaded cells differ from primary estimand. Alternative estimand 1 changes the strategy for selected clinical intercurrent events to "composite," resulting in an endpoint that comprises radiographic and clinical progression (as well as death). Alternative estimand 2 changes the strategy to "composite" only for systemic non-protocol anti-cancer medications, yielding an endpoint that comprises radiographic progression, death, or initiation of systemic NPACT.

Estimand Attribute [1]	**Alternative 1 Definition**		**Alternative 2 Definition**	
Population	Subjects randomized into the study intended to include patients with advanced HCC who have not received previous systemic anticancer therapy.		Subjects randomized into the study intended to include patients with advanced HCC who have not received previous systemic anticancer therapy.	
Endpoint	Duration of radiographic and clinical progression-free survival		Time to radiographic progression, death, or initiation of systemic NPACT	
Intercurrent events	**Event**	**Strategy**	**Event**	**Strategy**
	Receipt of non-assigned study treatment or non-receipt of study treatment	Treatment policy	Receipt of assigned study treatment	Treatment policy
	Clinical deterioration	Composite	Clinical deterioration	Treatment policy
	Receipt of local radiation to bone	Treatment policy	Receipt of local radiation to bone	Treatment policy
	Surgical resection of non-target tumor lesions	Treatment policy	Surgical resection of non-target tumor lesions	Treatment policy
	Surgical resection of target tumor lesions	Composite	Surgical resection of target tumor lesions	Hypothetical
	Receipt of systemic non-protocol anti-cancer medications	Composite	Receipt of systemic non-protocol anti-cancer medications	Composite
	Receipt of local non-protocol anti-cancer medications	Composite	Receipt of local non-protocol anti-cancer medications	Hypothetical
	Receipt of local radiation to soft tissue for disease under study	Composite	Receipt of local radiation to soft tissue for disease under study	Hypothetical
	Missing pre-randomization tumor assessment	Hypothetical	Missing pre-randomization tumor assessment	Hypothetical
	Missed two or more consecutive adequate tumor assessments immediately prior to rPD or death	Hypothetical	Missed two or more consecutive adequate tumor assessments	Hypothetical
Population summary	Difference in survival functions between treatment conditions.		Difference in survival functions between treatment conditions.	

1. Shaded cells differ from primary estimand.

[1]See Appendix E for estimand terminology

7.1.1.4 Primary Definition

For the primary analysis directed toward the primary estimand, duration of PFS is defined as the time from randomization to the earlier of either the date of radiographic progression per BIRC or the date of death due to any cause.

PFS (months) = (earliest date of progression, death, censoring – date of randomization + 1)/30.4375

General censoring rules for the primary analysis of PFS are described below, with details provided in Table 7Table 8 (analysis ID PFS-EP-1):

- Subjects who receive systemic or local NPACT or non-protocol radiation therapy (other than to bone) or surgery to resect target lesions before experiencing an event will be right censored at the date of the last tumor assessment prior to the date of initiation of subsequent therapy/surgery. If there is no such tumor assessment post randomization, the subject will be right censored on the date of randomization.

- Subjects who have not experienced an event (and are not otherwise censored) at the time of data cutoff will be right censored on the date of their last tumor assessment post randomization that is on or prior to the data cutoff date. If there is no such tumor assessment post randomization, the subject will be right censored on the date of randomization.

- Subjects who miss 2 or more scheduled tumor assessments followed by an event will be right censored on the date of their most-recent tumor assessment prior to the missing assessments. If there is no such tumor assessment post randomization, the subject will be right censored on the date of randomization.

7.1.1.5 Hypothesis

The hypotheses to be evaluated in the analysis of the PFS are:

H_0: $S(t)_{cabozantinib+atezolizumab} = S(t)_{sorafenib}$

H_A: $S(t)_{cabozantinib+atezolizumab} \neq S(t)_{sorafenib}$

where $S(t)_{cabozantinib+atezolizumab}$ and $S(t)_{sorafenib}$ are the survivor functions for PFS for the experimental and control arms, respectively.

7.1.1.6 Primary Analysis

The timing of this analysis is event-driven, and it will be conducted after at least 257 events have been observed in the PITT population. It is designed to include progression events as determined by the BIRC per RECIST 1.1. Clinical deterioration determined by the investigator will not be considered progression events. The actual number of events at the time of analysis may be higher than 257 due to the logistics of estimating event counts in an ongoing multicenter clinical study and predicting cutoff dates for analysis.

The hypothesis testing between experimental arm (cabozantinib + atezolizumab) compared to the control arm (sorafenib) will be performed using the stratified log-rank test with a 2-sided α=0.01 level of significance. The stratification factors are as described in Section 2.4 and the values used for analysis will be those recorded in the IRT.

The median duration of PFS along with the associated 99% and 95% confidence intervals (CI) for each study treatment arm will be estimated using the Kaplan-Meier method. The stratified hazard ratio (HR) and its 99% and 95% CI will be estimated using a Cox proportional-hazard model with treatment group as the independent variable using the same stratification factors as were used for the log-rank test.

If the p-value for the stratified log-rank test is less than the critical value for rejecting the null hypothesis and the HR ($\lambda_{\text{cabozantinib + atezolizumab}}/\lambda_{\text{sorafenib}}$) is < 1, the null hypothesis of no difference between the two study treatment arms will be rejected and it will be inferred that duration of PFS is greater in the experimental arm (cabozantinib + atezolizumab) compared to the control arm (sorafenib).

The unstratified versions of all the above analyses will also be conducted.

7.1.1.7 Sensitivity and Supplemental Analyses

Event and censoring definitions for the primary analysis (PFS-EP-1) and 3 sensitivity analyses (PFS-EP-2, PFS-EP-3, PFS-EP-4) directed at the primary estimated for PFS are provided in Table 8.

The three sensitivity analyses evaluate the impact of different assumptions or conditions that potentially influence the estimate of the primary estimand:

- The PFS-EP-2 definition evaluates the influence of potentially inconsistent tumor assessment intervals between arms. For subjects who experience radiographic

progression, it assigns the date of the scheduled visit as the event date, rather than the date of recorded progression.

- The PFS-EP-3 definition evaluates the influence of the assessor of radiographic progression and is based on RECIST 1.1 evaluations by the investigator rather than the BIRC.
- The PFS-EP-4 definition evaluates the influence of the missing tumor assessments. It classifies subjects who experience ≥ 2 consecutive missing scheduled ATA immediately prior to documented radiographic progression as having an event, rather than being censored, at the date of the last ATA prior to the missing visits.

Event and censoring definitions for 3 supplemental analyses (PFS-EA1-1, PFS-EA2-1, PFS-EA2-1) directed at the two alternative estimands for PFS are provided in Table 9.

- The PFS-EA1-1 definition is the primary analysis of alternative estimand 1.
- The PFS-EA2-1 definition is the primary analysis of alternative estimand 2.
- The PFS-EA2-2 definition is a sensitivity analysis of alternative estimand 2, similar to PFS-EP-4 (defined above).

Four additional "differential" sensitivity analyses (PFS-EP-11, PFS-EP-12, PFS-EP-13, PFS-EP-14) directed at the primary estimand will be conducted to evaluate the impact of potentially informative censoring. These analyses are based on the primary analysis PFS-EP-1 but with selected censored subjects re-classified as events, differentially by treatment arm, as shown in Table 10. These are highly conservative definitions intended to evaluate potential bias. The directions of treatment effects are of key interest rather than the values of inferential statistics.

The concordance in assessment of radiographic progression and the date between BIRC and investigator will be summarized.

All sensitivity and supplemental analyses will include all subjects in the PITT population. Tabulated summaries of survival times, hazard ratios, and log rank test statistics as well as graphs of survival functions will be presented.

Table 8: Event and Censoring Rules for Primary and Sensitivity Analyses of Primary PFS Estimand

Estimand	Primary		Primary		Primary		Primary	
Analysis type	Primary		Sensitivity		Sensitivity		Sensitivity	
Analysis purpose	Primary		Evaluate assessment time bias		Evaluate assessor bias		Evaluate potentially informative censoring	
Analysis ID	PFS-EP-1		PFS-EP-2		PFS-EP-3		PFS-EP-4	
Analysis name	rPFS per BIRC		Uniform dates		rPFS per Investigator		rPFS per BIRC despite missing ATA	
Estimand endpoint	Radiographic PD		Radiographic PD		Radiographic PD		Radiographic PD	
Population	PITT		PITT		PITT		PITT	
Situation	**Outcome**	**Date of Outcome**	**Outcome**	**Date of Outcome**	**Outcome**	**Date of Outcome**	**Outcome**	**Date of Outcome**
Radiographic PD per RECIST 1.1 per BIRC	event	date of recorded PD	event	date of scheduled visit (or next scheduled visit if between visits)	NA	NA	event	date of recorded PD
Radiographic PD per RECIST 1.1 per Investigator	NA	NA	NA	NA	event	date of recorded PD	NA	NA
Death	event	date of death	event	date of death	event	date of death	event	date of death
	Intercurrent events (excluding those with Treatment Policy strategy for all estimands)							
Clinical deterioration	NA	NA	NA	NA	NA	NA	NA	NA
Systemic NPACT (medications)	censored	date of last ATA* before first initiation of therapy	censored	date of last ATA* before first initiation of therapy	censored	date of last ATA* before first initiation of therapy	censored	date of last ATA* before first initiation of therapy
Local NPACT (medications)	censored	date of last ATA* before first initiation of therapy	censored	date of last ATA* before first initiation of therapy	censored	date of last ATA* before first initiation of therapy	censored	date of last ATA* before first initiation of therapy
Surgical resection of target tumor lesion(s)	censored	date of last ATA* before target lesion resection	censored	date of last ATA* before target lesion resection	censored	date of last ATA* before target lesion resection	censored	date of last ATA* before target lesion resection
Local radiation: to soft tissue for disease under study	censored	date of last ATA* before non-target lesion resection	censored	date of last ATA* before non-target lesion resection	censored	date of last ATA* before non-target lesion resection	censored	date of last ATA* before non-target lesion resection
	Missing data							
No baseline ATA	censored	date of rand.	censored	date of rand.	censored	date of rand.	censored	date of rand.
≥ 2 consecutive missing scheduled ATA immediately prior to analysis event	censored	date of last ATA* before missing visits	censored	date of last ATA* before missing visits	censored	date of last ATA* before missing visits	event	date of last ATA* before missing visits
	Observation ongoing							
None of the above	censored	date of last ATA*	censored	date of last ATA*	censored	date of last ATA*	censored	date of last ATA*

ATA, adequate tumor assessment; BIRC, blinded independent review committee; ITT, intent-to-treat; NA, not applicable; PD, progressive disease; NPACT, non-protocol anti-cancer therapy (medications including radiopharmaceuticals but excluding local radiation)

Date of recorded PD = see SAP text.
* or date of randomization if no post-randomization ATA
Blue cells indicate changes from primary analysis.

Table 9: Event and Censoring Rules for Supplementary Analyses of Alternative PFS Estimands

Estimand	Alternative 1		Alternative 2		Alternative 2	
Analysis type	Supplementary		Supplementary		Supplementary	
Analysis purpose	Alternate progression definition		Alternate progression definition		Evaluate potentially informative censoring	
Analysis ID	PFS-EA1-1		PFS-EA2-1		PFS-EA2-2	
Analysis name	Investigator claims		rPFS or sNPACT		rPFS or sNPACT despite missing ATA	
Estimand endpoint	Radiographic or clinical PD		Radiographic PD or initiation of systemic NPACT		Radiographic PD or initiation of systemic NPACT	
Population	PITT		PITT		PITT	
Situation	**Outcome**	**Date of Outcome**	**Outcome**	**Date of Outcome**	**Outcome**	**Date of Outcome**
Radiographic PD per RECIST 1.1 per BIRC	NA	NA	event	date of recorded PD	event	date of recorded PD
Radiographic PD per RECIST 1.1 per Investigator	event	date of recorded PD	NA	NA	NA	NA
Death	event	date of death	event	date of death	event	date of death
	Intercurrent events (excluding those with Treatment Policy strategy for all estimands)					
Clinical deterioration	event	date of determination of clinical deterioration	NA	NA	NA	NA
Systemic NPACT (medications)	event	date of first initiation of therapy	event	date of first initiation of therapy	event	date of first initiation of therapy
Local NPACT (medications)	event	date of first initiation of therapy	censored	date of last ATA* before first initiation of therapy	censored	date of last ATA* before first initiation of therapy
Surgical resection of target tumor lesion(s)	event	date of last ATA* before target lesion resection	censored	date of last ATA* before target lesion resection	censored	date of last ATA* before target lesion resection
Local radiation: to soft tissue for disease under study	event	date of last ATA* before non-target lesion resection	censored	date of last ATA* before non-target lesion resection	censored	date of last ATA* before non-target lesion resection
	Missing data					
No baseline ATA	censored	date of rand.	censored	date of rand.	censored	date of rand.
≥ 2 consecutive missing scheduled ATA immediately prior to analysis event	censored	date of last ATA* before missing visits	censored	date of last ATA* before missing visits	event	date of last ATA* before missing visits
	Observation ongoing					
None of the above	censored	date of last ATA*	censored	date of last ATA*	censored	date of last ATA*

ATA, adequate tumor assessment; BIRC, blinded independent review committee; ITT, intent-to-treat; NA, not applicable; PD, progressive disease; NPACT, non-protocol anti-cancer therapy (medications including radiopharmaceuticals but excluding local radiation)
Date of recorded PD = see SAP text.
* or date of randomization if no post-randomization ATA
Blue cells indicate changes from primary analysis.

Table 10: Definitions for Differential Sensitivity Analyses of PFS to Further Evaluate Potentially Informative Censoring (based on PFS-EP-1)

Estimand	Primary							
Analysis type	Sensitivity							
Analysis purpose	Differential analyses to explore potentially informative censoring							
Analysis strategy	Selected censored subjects in primary analysis reclassified as events, differentially by assigned treatment arm, as shown below (green cells)							
Analysis ID	**PFS-EP-11**		**PFS-EP-12**		**PFS-EP-13**		**PFS-EP-14**	
Censoring category for subjects censored in PFS-EP-1 in PITT Population	**Exp**	**Cntrl**	**Exp**	**Cntrl**	**Exp**	**Cntrl**	**Exp**	**Cntrl**
rPD per INV but not BIRC	At recorded date of first rPD per INV		At recorded date of first rPD per INV		At recorded date of first rPD per INV	At recorded date of first rPD per INV	At recorded date of first rPD per INV	
Discontinued radiographic assessment for reasons other than rPD and no sNPACT	At date of last ATA		At date of last ATA		At date of last ATA		At date of last ATA	
sNPACT prior to rPD per INV or BIRC			At date of first sNPACT	At date of first sNPACT			At date of first sNPACT	

ATA, adequate tumor assessment; BIRC, blinded independent review committee; Cntrl, control arm; Exp, experimental arm; rPD, radiographic progressive disease per RECIST 1.1;
INV, investigator; sNPACT, systemic non-protocol anti-cancer therapy (medications including radiopharmaceuticals)
* or date of randomization if no post-randomization ATA

7.1.1.8 Exploratory Analyses

Exploratory analyses of the effect of baseline characteristics, stratification factors, and other variables on PFS may be conducted if necessary, using Cox regression models and subgroup analyses performed employing Kaplan-Meier methods. At the time of the primary analysis, PFS-EP-1 will also be repeated in the full ITT population.

The concordance in assessment of radiographic progression and the date between BIRC and investigator will also be summarized for the ITT population.

7.1.2 Duration of Overall Survival (OS)

7.1.2.1 Primary Estimand

The primary estimand for OS is difference in survival functions between treatment conditions in the duration of overall survival in the targeted patient population, irrespective of whether the assigned study treatment was given and irrespective of whether non-protocol anti-cancer therapy or radiation of any kind was given.

Derived as follows:

Estimand Attribute [1]	**Primary Definition**		**Rationale (as needed)**
Population	Subjects randomized into the study intended to include patients with advanced hepatocellular carcinoma who have not received previous systemic anticancer therapy.		The initial clause "Subjects randomized into the study intended to include…" is included to align the estimand with the application of the ITT principle expected by regulatory reviewers.
Endpoint	Duration of overall survival		
Intercurrent event(s)	**Event**	**Strategy**	**Rationale (as needed)**
	Receipt of non-assigned study treatment or non-receipt of study treatment	Treatment policy	This strategy applied to align the estimand with the application of the ITT principle expected by regulatory reviewers.
	Receipt of non-protocol anti-cancer therapy or radiation of any kind	Treatment policy	
	Lost to follow-up	Hypothetical	This strategy arises as a consequence of the convention of censoring at loss to follow-up and the assumption in the Kaplan-Meier model that censored subjects would have behaved in a fashion similar to those not censored, i.e. not experienced these intercurrent event(s)
Population summary	Difference in survival functions between treatment conditions		

[1]See Appendix E for estimand terminology

7.1.2.2 Alternative Estimand

One alternative estimands for OS is defined as below, arising from changes in strategy for handling some intercurrent events to "hypothetical." This analysis is intended to "factor out" the

effects of non-protocol anti-cancer therapy on overall survival. Shaded cells differ from primary estimand.

Estimand Attribute [1]	**Alternate definition**	
Population	Subjects randomized into the study intended to include patients with advanced hepatocellular carcinoma who have not received previous systemic anticancer therapy.	
Endpoint	Duration of overall survival	
Intercurrent event(s)	**Event**	**Strategy**
	Receipt of assigned study treatment	Treatment policy
	Receipt of systemic non-protocol anti-cancer medications	Hypothetical
	Receipt of local liver-directed non-protocol anti-cancer medications	Hypothetical
	Lost to follow-up	Hypothetical
Population summary	Difference in survival functions between treatment conditions	

[1]See Appendix E for estimand terminology

7.1.2.3 Primary Definition

For the primary analysis directed toward the primary estimand (analysis ID OS-EP-1), duration of OS is defined as the time from randomization to death due to any cause. For subjects who are not known to have died at the time of data cutoff and are permanently lost to follow-up, duration of OS will be right censored at the earlier of the following dates: date the subject was last known to be alive or date of full withdrawal of consent (including survival follow-up), or date of data cutoff.

OS (months) = (earlier of date of death or censoring – date of randomization + 1)/30.4375

7.1.2.4 Hypothesis

The hypotheses to be evaluated in the analysis of the OS are:

H_0: $S(t)_{\text{cabozantinib + atezolizumab}} = S(t)_{\text{sorafenib}}$

H_A: $S(t)_{\text{cabozantinib + atezolizumab}} \neq S(t)_{\text{sorafenib}}$

where $S(t)_{cabozantinib+atezolizumab}$ and $S(t)_{sorafenib}$ are the survivor functions for OS for the experimental and control arms, respectively.

7.1.2.5 Primary Analysis

The timing of this analysis is event-driven and it will be conducted after at least 368 deaths (overall survival events) have been observed in the experimental and control arms in the ITT population. The actual number of events at the time of analysis may be higher than 368 due to

the logistics of estimating event counts in an ongoing multicenter clinical study and predicting cutoff dates for analysis.

The hypothesis testing between the experimental arm (cabozantinib + atezolizumab) compared to the control arm (sorafenib) will be performed using a 2-sided stratified log-rank test. Please see Section 7.2 for details regarding the allocation of overall Type I error (α-level). The stratification factors are as described in Section 2.4 and the values used for analysis will be those recorded in the IRT.

Additionally, up to two interim analyses of OS are planned at approximately the 33% (conditional on positive PFS endpoint inference) and 66% information fractions. The critical p-values (and observed HR) for testing these interim hypotheses will be based on Lan-DeMets O'Brien-Fleming alpha-spending function. The overall α to be spent on testing the duration of OS via the use of a spending function will as described in Section 7.2. The actual critical values for each OS analysis will depend upon the true number of events observed at conduct of each analysis. Details are provided in Section 7.3.

The median duration of OS along with the associated (1-α) % and 95% confidence intervals (CI) for each study treatment arm will be estimated using the Kaplan-Meier method. The stratified hazard ratio (HR) along with the associated (1-α) % and 95% CIs will be estimated using a Cox proportional-hazard model with treatment group as the independent variable using the same stratification factors as were used for the log-rank test.

If at any OS analysis (interim or primary) the p-value for the stratified log-rank test is less than the critical value for rejecting the null hypothesis and the HR (λ cabozantinib + atezolizumab/ λ sorafenib) is < 1, the null hypothesis of no difference between the two study treatment arms will be rejected and it will be inferred that duration of OS is greater in the experimental arm (cabozantinib + atezolizumab) compared to the control arm (sorafenib).

The unstratified versions of all the above analyses will also be conducted.

7.1.2.6 Supportive Analyses

A supportive analysis (OS-EA1-1) directed at the alternative estimand will be conducted in the experimental and control arms in the ITT population. In this analysis a subject receiving non-protocol systemic or local-liver directed anticancer therapy (medications) will be right censored at the first start date of such subsequent anticancer therapy. Summaries will be provided as for

the primary analysis of OS.

7.1.2.7 Exploratory Analyses

Exploratory analyses of the effect of baseline characteristics, stratification factors, and other variables on OS may be conducted if necessary, using Cox regression models and subgroup analyses performed employing Kaplan-Meier methods.

An exploratory analysis of OS between the single-agent cabozantinib and sorafenib arms in the ITT population will also be performed following the same methodology as in Section 7.1.2.5.

7.2 Control of Type I Error

Inflation of Type 1 error associated with testing of multiple-primary endpoints and a secondary efficacy endpoint (PFS between single-agent cabozantinib and control) will be controlled by a modified Bonferroni procedure and a parallel gatekeeping method (FDA 2017). Initially, the study-wise 2-sided alpha of 5% will be unequally allocated between primary endpoints PFS (1%) and OS (4%). The secondary PFS endpoint will be tested at the 1% level only if the primary PFS test is successful. If the null hypotheses are rejected for both the primary and secondary PFS endpoints, the 1% alpha for PFS will be reallocated to OS, allowing it to be tested at the 5% level.

Two interim analyses of OS are planned at the 33% and 66% information fractions (the first interim will coincide with the primary analysis of PFS and will be performed only if the null hypothesis for PFS is rejected analysis and is expected to occur at approximately the 33% information fraction). Type 1 error associated with interim analyses will be controlled using Lan-DeMets O'Brien-Fleming (LD-OF) alpha-spending functions based upon a 4% total alpha allocation for OS. However, if the null hypotheses are rejected for both the primary and secondary PFS endpoints, the 1% alpha for PFS will be reallocated to OS per the fallback method. As a consequence, the final OS analysis will employ critical values from the spending function at 5% level. The alpha level of the interim analysis at 66% will be dependent on the timing of the interim analysis. It can be conducted at 5% alpha level only if it occurs after the null hypothesis for the secondary PFS is rejected. (see Section 7.3).

All other statistical evaluations of efficacy will be considered exploratory.

7.3 Interim Analyses

The number of events required to evaluate OS is based upon assumptions currently available and provides high power to detect the smallest clinically meaningful difference in OS under these assumptions. However, as there is uncertainty in the assumptions, interim analyses provide an opportunity to stop the trial early if the treatment benefit of the experimental arm is larger than expected, potentially allowing the new regimen to become available sooner to this patient population.

Two interim analyses (IA) of OS are planned at approximately the 33% (IA1) and 66% (IA2) information fractions based on a total alpha of 4%. However, if the null hypotheses are rejected for both the primary and secondary PFS endpoints, the 1% alpha for PFS will be reallocated to OS per the fallback method (FDA Guidance for Industry: Multiple Endpoints in Clinical Trials (draft), 2017). As a consequence, the OS analyses may employ critical values from the spending function at the 5% level if performed after the null hypothesis for the secondary PFS is rejected.

Table 11: Details and Boundaries for the Interim and Final Analyses

	Critical Values for Testing OS			
OS alpha total	4.0%		5.0%[a]	
OS analysis information fraction	Critical p-value	HR to reject	Critical p-value	HR to reject
33%[b] (IA1)	≤ 0.00011	≤ 0.478	≤0.00021[b]	≤ 0.482[b]
66% (IA2)	≤ 0.00867	≤ 0.701	≤ 0.01202	≤ 0.705
100% (final)	≤ 0.03728	≤ 0.794	≤ 0.04626	≤ 0.797

HR, hazard ratio; NA, not applicable; OS, overall survival.

[a] Analyses of OS will be conducted at 5% alpha level only if the null hypotheses for both the primary and secondary PFS endpoints are rejected prior to testing OS.

[b] OS will not be tested at the 33% information fraction unless the null hypothesis for PFS has been rejected.

Rejection of the null hypotheses for OS at the first interim analysis at about 33% information is not expected: it is designed primarily as an administrative analysis to coincide with the primary analysis of PFS and will be performed only if the null hypothesis for PFS is rejected.

Due to logistical considerations in event ascertainment and operational planning and conduct, the actual analyses may include more or fewer events than the target information fractions. The

actual critical values employed at the interim and final analyses of OS will depend upon the actual information fraction at the time of the analyses.

Should the timing of the first interim analysis of OS (IA1) trend towards an information fraction much higher than expected (e.g. due to faster-than-expected deaths or slower-than expected PFS events), the interim analyses of OS will be conducted as follows.

OS IA1 IF	Proceed as	Rationale
<50%	perform IA1 and IA2 as planned	similar conditions to original plan
50 to <66%	• perform IA1 at time of primary PFS, irrespective of whether PFS null hypothesis is rejected or not • Conduct IA2 to 80% IF (instead of 66%)	• Higher than expected information for OS negates administrative nature of the analysis. • Preserves two IA as planned with reasonable intervals between them
≥66%	perform a single IA at 66% IF, irrespective of timing or results of primary PFS analysis.	Information fraction meets or exceeds that of planned analysis at 66%

An interim analysis of secondary PFS is planned contemporaneously with the primary PFS analysis and will be conducted only if the null hypothesis for primary PFS is rejected. It is anticipated to be at approximately 67% information fraction based on a total alpha of 1%.

If null hypothesis is rejected for a given endpoint at interim analysis (secondary PFS or OS), no subsequent testing of that endpoint will be performed.

Interim analysis of primary PFS and futility analyses of PFS or OS are not planned.

7.4 Secondary Efficacy Endpoint

7.4.1 Progression Free Survival (PFS) per BIRC between Single-Agent Cabozantinib and Control Arms

The secondary efficacy endpoint for this study is PFS per BIRC between single-agent cabozantinib and the control arms in the ITT population. The timing of this analysis is event-driven, and it will be conducted after at least 283 events have been observed in the two arms. Formal hypothesis testing is planned for this endpoint following the same methodology as outlined in Section 7.1.1.6. One interim analysis of secondary PFS is planned, to be

contemporaneous with the primary PFS analysis, at approximately the 67% information fraction (Section 7.3)

A supportive analysis of PFS per INV will also be performed on the ITT population.

7.5 Additional Endpoints

Each exploratory endpoint will be analyzed using an appropriate two-sided statistical test without adjustment for multiplicity unless specified otherwise. Statistical results for exploratory endpoints will be considered supportive. Exploratory analyses will be performed on the PITT and ITT populations unless specified otherwise.

7.5.1 Objective Response Rate (ORR) by BIRC

For each subject, best overall response (BOR) is defined as the best tumor assessment category as determined per RECIST 1.1 by BIRC that occurs through the first overall time point response of PD and prior to any of the censoring events defined for the primary analysis of PFS as described in Section 7.1.1. Tumor assessment categories are ranked as: confirmed complete response (CR), confirmed partial response (PR), stable disease (SD), progressive disease (PD) and not evaluable (NE). To be classified as confirmed CR or confirmed PR, confirmation must have occurred on a subsequent visit that is ≥ 28 days after the response was first observed. To be classified as SD, at least one overall time point response of SD must be documented ≥ 21 days after randomization. The ORR is defined as the proportion of subjects with a BOR of confirmed CR or confirmed PR.

Point estimates of ORR for each study treatment arm, the difference in ORR between the two study treatment arms and associated (1-α) % confidence intervals will be provided. The odds ratio and associated (1-α) % confidence intervals will also be presented. The 2-sided CIs for the point estimate will be calculated using exact methods. The 2-sided CIs for the difference in ORR between the two study treatment arms and for the odds ratio will be calculated using asymptotic methods. All summaries for ORR will be provided for the PITT and ITT Populations

Waterfall plots displaying maximum percent tumor reduction since baseline in target lesions will be generated for tumor assessment data per BIRC.

These plots will include subjects with a tumor assessment at baseline and at least one post-baseline visit. For each subject, data on or after the progression/censoring date of the respective PFS analyses described in Section 7.1.1 will be excluded from the waterfall plots.

7.5.2 ORR by Investigator

The ORR per RECIST 1.1 by investigator will be analyzed as described in Section 7.5.1 above, respectively. Only 95% confidence levels will be used in these exploratory analyses. The P-ITT and ITT populations will be used.

7.5.3 Time to Progression (TTP)

Time to progression is defined as the time from randomization until tumor progression per RECIST v1.1. Subjects who die due to progressive disease will be considered as events and those who die for other reasons will be censored on the day they died.

Time to progression (months) = (earliest date of radiographic progressive disease or death due to progressive disease or censoring – date of randomization + 1)/30.4375

The TTP will be analyzed separately for radiographic tumor assessments per BIRC and per Investigator for both PITT and ITT populations. For TTP, the dates of progression and censoring per BIRC are shown in column PFS-EP-1 and PFS-EP-3 of Table 8 respectively. However, only deaths due to progressive diseases will be considered as events and those due to other reasons will be censored at the date of death.

Time to progression will be analyzed using the same analysis methods (Kaplan-Meier and Cox proportional hazards model) as for the analysis on PFS (see Section 7.1.1.6).

7.5.4 Duration of Objective Response (DOR)

Duration of objective response is defined as the time from the first documentation of objective response that is subsequently confirmed at a visit that is ≥ 28 days later to disease progression or death due to any cause.

Duration of response (months) = (earliest date of progressive disease or death due to any cause or censoring – date of first objective response + 1)/30.4375

Duration of objective response will be computed only among subjects who experience an objective response (confirmed CR or confirmed PR). DOR will be computed both by BIRC and by Investigator and analyzed separately. For DOR per BIRC and per Investigator, the dates of progression and censoring are shown in column PFS-EP-1 and PFS-EP-3 of Table 8, respectively.

Duration of objective response will be analyzed using the same analysis method (Kaplan-Meier) as for the analysis on PFS (see Section 7.1.1.6) for both P-ITT and ITT populations.

7.5.5 PFS and ORR per modified RECIST (mRECIST)

The analysis of PFS and ORR per BIRC using modified RECIST (Lencioni, 2010) criteria will be as described in Sections 7.1.1.3 and 7.4 above, respectively. Modified RECIST measurements will be derived computationally from reader-approved target lesion contours and subjective non-target lesion assessments by the BIRC based upon standard CT/MRI images obtained for the RECIST 1.1 assessments. Only 95% confidence levels will be used in these exploratory analyses. Summary tables will be provided for both the PITT and ITT population.

7.5.6 Pharmacokinetics (PK)

Blood samples will be obtained from all subjects in the experimental arm (cabozantinib + atezolizumab) and the single-agent cabozantinib arm for cabozantinib concentration measurements. Blood samples will be obtained from all subjects in the experimental arm for atezolizumab concentration measurement. Pharmacokinetics analyses of cabozantinib and atezolizumab are outside the scope of this plan. A separate PK analysis plan and PK report will be provided. In addition, population PK and exposure-response analyses of cabozantinib will be explored and reported separately. PK analyses will also be conducted among the subset of subjects enrolled in China.

7.5.7 Immunogenicity Assessments

This is beyond the scope of this plan.

7.5.8 Serum Alpha-Fetoprotein (AFP)

Serum AFP is assessed in the study at various time points such as baseline, W4D1, W7D1, W10D1, W13D1, FU1, and FU2. Serum AFP will be analyzed as a continuous and discrete variable (<400 ng/mL, ≥400 ng/mL) to assess its potential as a prognostic marker. The serum levels and change from baseline (absolute and percent change) will be summarized using descriptive statistics at various measurement time points and presented by study treatment arm for both PITT and ITT populations. Descriptive statistics for best/worst percent change since baseline will also be presented per arm using all available data at protocol-defined time points.

Best change is defined as the largest reduction (or smallest increase if no reduction) and worst change is defined as the largest increase (or smallest reduction if no increase).

Appropriate transformations may be applied to normalize the data for presentation or analysis. AFP may also be categorized based on previously published measures. For example, a decrease from baseline of 20% or more in serum AFP by or at a specific time point (say, W7D1) may be defined post-hoc as a prognostic response.

Exploratory modeling analyses including survival analyses with time dependent covariates, general linear models, and mixed models with repeated measures will be employed to assess meaningful relationships between AFP levels or changes from baseline and clinical outcomes.

A waterfall plot of best percentage change (i.e., best percentage decrease) from baseline in AFP will be presented for each treatment group using the PITT and ITT populations. In all calculations, best change from baseline will be based only on protocol-defined time-points.

7.5.9 Peripheral Biomarkers

The tumor PD-L1 status will be used for subgroup analysis as discussed in Section 7.6. Additional peripheral biomarker analysis is outside the scope of this plan.

7.5.10 Health-Related Quality of Life (HRQOL)

Health-related quality of life (HRQOL) will be assessed by the EuroQol Health questionnaire instrument (EQ-5D-5L) (Reenen MV, 2019). The questionnaire will be self-completed by the subjects at various time points until disease progression and will provide a generic measure of health for clinical appraisal (see protocol Section 5.7.7). The EQ-5D-5L questionnaire has two pages: a descriptive page which assesses on an increasing severity scale of 1-5 changes in the following five questions (dimensions): mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. The second page has a 0-100 visual analogue scale (VAS) which records the subjects self-rated health between 100 (best health you can imagine) and 0 (worst health you can imagine) and serves as quantitative measure of health by the subject (see protocol Appendix D and EQ-5D-5L User Guide 2019).

To compare each of the cabozantinib containing arms with sorafenib, the following summaries are planned at each time point for each of the 6 questions for PITT and ITT populations:

Within each treatment arm:

- Descriptive statistics (number of observations, mean and standard deviation)
- Rate of completion for the questionnaire at each time point. This is defined as total number of subjects who answered all questions on the EQ-5D-5L questionnaire / the expected total number of subjects still on study at the visit

- Mean change from baseline at each time point and the corresponding 95% CI and p-value from one-sample t-test
- Effect size for change from baseline within arm, calculated as (mean of change in score/pooled standard deviation for baseline scores). An effect size greater than ≥ 0.3 will be considered potentially clinically meaningful
- Shift in the severity scale since baseline

Across treatment arms

- The difference in the effect sizes will be presented
- Difference in mean change from baseline for the questions at the time point will be evaluated by a two-sample t-test
- Line plots for mean ± standard error and the corresponding mean for change from baseline over time. Data from all treatment arms will be displayed on the same plot. In addition, these plots will also show the average state of the subjects at 3 landmark points, namely, around end of treatment, around progression and around 30 days post treatment follow-up for the two treatment arms
- Percentage of subjects in Level 1 (no problem) vs. Levels 2-5 (any problems) will be summarized over time
- Percentage of subjects with any problems (Level 2-5) will be compared between the treatment arms using a bar chart

The EQ-5D-5L may be converted into a single index (EQ-Index) value normalized across different countries where the index is validated. See Appendix D for conversion details. For EQ-VAS and EQ-Index, descriptive statistics for change from baseline at each time will be presented. Plots for mean ± standard error and mean change from baseline ± standard error over all time points for the two treatment arms will be generated. In addition, repeated-measures mixed-effects models will be used to explore treatment differences over time. These analyses will include the outcome variable of QOL score change from baseline. The predictors (fixed effects) will be the baseline scores, treatment arms, visit, and randomization strata described in Section 2.4. The individual subject nested within the planned treatment arm will be the random effect. All available data will be included for the analysis. The estimated least squares means for the two treatment arm and their difference, the p-values comparing the 2 treatment arms and the effect size will be presented. No adjustment will be made for multiple comparisons. An effect size of difference ≥ 0.3 will be considered potentially clinically meaningful.

7.5.11 Health Care Resource Utilization

For this study the following health care resource utilization (HCRU) parameters collected during the study observation period will be summarized for the Safety population:

- Type of admission (hospital, intensive care unit (ICU), emergency unit)
- Descriptive statistics of number of hospital visits per subject
- Descriptive statistics of number of ICU visits per subject
- Descriptive statistics of number of emergency room (ER) visits per subject
- Total person-years of observation
- Total days of hospitalization
- Days of hospitalization per person-year
- Total days of ICU visit
- Days of ICU visit per person-year

Person-years total is calculated as the sum of [(end date of safety observation period – first dose date +1)/365.25] across all subjects in the given treatment group. To calculate the per person year value for a subject for a HCRU parameter, the numerator is the sum of the days or visits for that subject for the parameter; and the denominator is the applicable person-years of observation as defined.

7.6 Subgroups

The following subgroups based on baseline characteristics and stratification factors as reported by the investigator on the CRF will be explored for primary and secondary efficacy endpoints in the PITT and ITT populations. Forest plots will be provided for hazard ratios/odds ratio as applicable.

- Age category
  - <65 years
  - 65 to <75 years
  - 75 to <85 years
  - ≥85 years
- Sex
  - Male
  - Female

- Race
  - Asian
  - Black or African American
  - White
  - Rest of the races reported/Not Reported
- Geographic Regions (not mutually exclusive)
  - Asia
    - China + Hong Kong + Taiwan
  - Other Region (not Asia)
- ECOG Performance status at baseline:
  - 0
  - 1
- Etiology of disease per stratification factors per CRF:
  - HBV [with or without HCV]
  - HCV [without HBV]
  - Other
- Current etiology of disease (per cancer history CRF):
  - HBV [without HCV] (Yes)
  - HCV [without HBV] (Yes)
  - HBV and HCV (Yes)
  - Alcoholism (Yes)
  - Nonalcoholic Steatohepatitis (NASH) (Yes)
  - Other (Yes)
- Presence of extrahepatic spread of disease and/or macrovascular invasion (per cancer history CRF):
  - Yes
  - No
- Macrovascular invasion
  - Yes
  - No
- Extrahepatic disease
  - Yes
  - No
- Cirrhosis
  - Yes
  - No

- Lung metastasis per investigator
    - No
    - Yes
- Important visceral metastasis per investigator
    - No
    - Yes
- Number of organs at screening per investigator
    - 1
    - 2
    - ≥3
- AFP at baseline
    - <400 ng/mL
    - ≥ 400 ng/L
- Tumor PD-L1 status
    - Positive
    - Negative
    - Unknown

8 SAFETY SUMMARIES

Safety summaries will include all subjects in the Safety and P-Safety populations. Data to be included for Core and Crossover Periods and the treatment arm groupings for the summaries are described in Section 4.6. No formal statistical comparison between the experimental and control arms is planned.

Safety and tolerability will be assessed by the incidence of treatment emergent-adverse events (TEAEs), changes in laboratory parameters and vital signs from baseline, and ECOG PS.

An Independent Data Monitoring Committee (IDMC) will monitor safety of the subjects during the study on a regular basis. The committee will operate independently from the Sponsor and the clinical investigators.

The primary responsibility of the IDMC is to review the accumulating safety data on a regular and an ad hoc basis and make recommendations to the Sponsor regarding the continued conduct of the study. Safety data will be provided at regular intervals to the IDMC in the form of summary reports or data listings from the Sponsor or its designated representative.

Details regarding IDMC membership, schedule and format of meetings, format for presentation of data, access to interim data, method, and timing of providing interim reports to the IDMC, and other issues relevant to committee operations are described in the IDMC charter.

The IDMC members will use their expertise, experience, and judgment to evaluate the safety data from the trial and to recommend to the Sponsor whether the trial should continue or be stopped early for safety. No formal statistical rules recommending early stopping for safety are planned.

8.1 Adverse Events

Adverse event terms recorded on the CRF will be mapped to preferred terms and system organ class using the Medical Dictionary for Regulatory Activities (MedDRA). The severity of AEs will be measured by CTCAE ((Common Terminology Criteria for Adverse Events, version 5.0, 2017)) guidelines. The investigator will judge each event to be "not related" or "related" to study treatment. For subjects in the experimental arm, the investigator will judge each event's relationship to study treatment, separately for cabozantinib and atezolizumab. However, in summarizing relationship of study treatment to adverse events in the experimental arm, relationship to either cabozantinib or atezolizumab will be considered as related to the experimental arm.

A TEAE is defined as any event with an onset date on or after the date of the first dose of study drug or any ongoing event on the date of the first dose of study drug that worsens in severity after the date of the first dose of study drug. In the Crossover phase, subjects who cross over to receive the experimental treatment, this definition will be with respect to the first dose of the experimental treatment received.

Unless otherwise specified, only TEAEs with an onset date (or date of increase in severity) through the end of the relevant safety observation period (see Section 4.6) will be tabulated in summary tables.

All deaths after informed consent, irrespective of when they occur, are classified as Grade 5 TEAEs. However, only those associated with an event starting after informed consent through 100 days after the date of the decision to discontinue the last component study treatment (or

starting at any time after informed consent and judged to be related to study treatment) required reporting as an SAE.

For the purpose of calculating treatment emergence and inclusion in summary tables, incomplete onset dates will be imputed as detailed in Appendix A. The calculations of percentages will be based on original unrounded values.

Summaries of AEs leading the changes in study treatment (e.g. leading to dose hold, reduction, modification, and/or discontinuation) will be based upon the fields for actions taken with each agent, collected on the AE CRF, e.g.

- Summaries of AEs leading to hold of oral treatment will include records where the "action taken" with cabozantinib or sorafenib = dose hold
- Summaries of AEs leading to hold of IV treatment will include records where the "action taken" with atezolizumab = dose delay
- Summaries of AEs leading to hold of ***any*** treatment will include records where the "action taken" with cabozantinib OR sorafenib is dose hold OR that for atezolizumab is dose delay
- Summaries of AEs leading to hold of ***all*** treatment will include records where the "action taken" with cabozantinib AND atezolizumab (for the combination regimen) or with sorafenib (for the control arm) = dose hold/delay as applicable

An overall summary of adverse events will be provided with the number and percent of subjects who experienced the following types of events during the standard safety observation period (unless otherwise noted as "at any time" or "during the extended safety observation periods" below) in each treatment arm:

- Subjects with a TEAE
- Subjects with a TEAE Related to Any Treatment Component
- Subjects with a Serious TEAE
- Subjects with a Serious TEAE during the extended safety observation periods
- Subjects with a Serious TEAE Related to Any Treatment Component at any time
- Subjects with a Worst-Grade 3 or 4 TEAE
- Subjects with a Worst-Grade 3 or 4 TEAE Related to Any Treatment Component
- Subjects with a Worst-Grade 4 TEAE
- Subjects with a Worst-Grade 4 TEAE Related to Any Treatment Component
- Subjects with a Grade 5 TEAE (death)

- Subjects with a Grade 5 TEAE (death) judged not to be causally related to disease under study
- Subjects with a Grade 5 TEAE Related to Any Treatment Component at any time (all treatment-related deaths)
- Subjects with a Grade 5 TEAE (death) Related to Any Treatment Component
- Subjects with a Grade 5 TEAE (death) during the extended safety observation period
- Subjects with a Grade 5 TEAE (death) judged not to be causally related to disease under study during the extended safety observation period
- Subjects with a Grade 5 TEAE (death) Related to Any Treatment Component during the extended safety observation period
- Subjects with a TEAE leading to Dose Modification
    - Modification of any treatment component
    - Modification of IV treatment
    - Modification of oral treatment
    - Modification of all treatment components
    - Modification of IV treatment only
- Modification of oral treatment only
- Subjects with a TEAE leading to Dose Reduction of oral treatment
- Subjects with a TEAE leading to Dose Hold of any treatment component
- Subjects with TEAE leading to Treatment Discontinuation at any time
        - Discontinuation of any treatment component
        - Discontinuation of IV treatment
        - Discontinuation of oral treatment
        - Discontinuation of all treatment components
        - Discontinuation of IV treatment only
        - Discontinuation of oral treatment only
    - TEAEs related to study treatment
        - Discontinuation of any treatment component
        - Discontinuation of IV treatment
        - Discontinuation of oral treatment
        - Discontinuation of all treatment components
        - Discontinuation of IV treatment only
        - Discontinuation of oral treatment only
    - TEAEs not related to any treatment component
    - TEAEs not causally related to disease under study
    - TEAEs causally related to disease under study

The following summaries of AEs will be provided:

TEAE included	Row-levels (sorted by)	Columns will Display	Safety Observation Period
Subject Incidence by SOC, Preferred Term and Severity			
All	SOC and PT (MedDRA standard)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
Related to any treatment component	SOC and PT (MedDRA standard)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
Serious	SOC and PT (MedDRA standard)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
Subject Incidence by Preferred Term and Severity			
All	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
	PT (descending frequency of Grade 3/4)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 1, Grade 2, Grade 3, Grade 4, Grade 5	Standard
Related to any treatment component	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
	PT (descending frequency of Grade 3/4)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
Serious	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Extended
Serious and Related to any treatment component	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	At any time
Leading to Treatment Modification			
Leading to delay or hold of any treatment component	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
	PT (descending frequency of Grade 3/4)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
Leading to delay or hold of all treatment components	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
	PT (descending frequency of Grade 3/4)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
Leading to hold of oral study treatment	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
	PT (descending frequency of Grade 3/4)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard

TEAE included	Row-levels (sorted by)	Columns will Display	Safety Observation Period
Leading to IV treatment delay	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
	PT (descending frequency of Grade 3/4)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
Leading to dose reduction of oral treatment	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
	PT (descending frequency of Grade 3/4)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
Leading to dose modification of any treatment component	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
	PT (descending frequency of Grade 3/4)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
Leading to dose modification of all treatment components	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
	PT (descending frequency of Grade 3/4)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
Leading to dose modification of oral treatment	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
	PT (descending frequency of Grade 3/4)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
Leading to dose modification of IV treatment	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
	PT (descending frequency of Grade 3/4)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
Leading to Treatment Discontinuation			
Leading to discontinuation of any treatment component	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	At any time
	PT (descending frequency of Grade 3/4)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	At any time
Leading to discontinuation of all study treatment components	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	At any time
	PT (descending frequency of Grade 3/4)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	At any time
Leading to discontinuation of oral treatment	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	At any time
Leading to discontinuation of IV treatment	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	At any time
Leading to discontinuation of any treatment component and related to study treatment	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	At any time
Leading to discontinuation of all treatment components and related to study treatment	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	At any time

TEAE included	Row-levels (sorted by)	Columns will Display	Safety Observation Period
Leading to discontinuation of oral treatment and related to study treatment	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	At any time
Leading to discontinuation of IV treatment and related to study treatment	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	At any time
Leading to discontinuation of any treatment component and not related to study treatment and not causally related to disease under study	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	At any time
Leading to discontinuation of all treatment components and not related to study treatment and not causally related to disease under study	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	At any time
Leading to discontinuation of oral study treatment and not related to study treatment and not causally related to disease under study	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	At any time
Leading to discontinuation of IV study treatment and not related to study treatment and not causally related to disease under study	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	At any time
Leading to discontinuation of any treatment component and not related to study treatment and causally related to disease under study	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	At any time
Leading to discontinuation of all treatment components and not related to study treatment and causally related to disease under study	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	At any time
Leading to discontinuation of oral study treatment and not related to study treatment and causally related to disease under study	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	At any time
Leading to discontinuation of IV study treatment and not related to study treatment and causally related to disease under study	PT (descending frequency of Any Grade)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	At any time
Subject Incidence of Grade 5 AEs with Odds ratio, Relative Risk and Risk Difference			
Grade 5	PT (descending frequency of difference)	All events	Standard
	PT (descending frequency of Grade 5)	Worst severity: Grade 5	Standard
	PT (descending frequency of Grade 5)	Worst severity: Grade 5	Extended
Grade 5 AEs judged not to be causally related to disease under study	PT (descending frequency of Grade 5)	Worst severity: Grade 5	Standard

TEAE included	Row-levels (sorted by)	Columns will Display	Safety Observation Period
	PT (descending frequency of Grade 5)	Worst severity: Grade 5	Extended
Grade 5 AE related to any treatment component	PT (descending frequency of Grade 5)	Worst severity: Grade 5	At any time
	PT (descending frequency of Grade 5)	Worst severity: Grade 5	Standard
	PT (descending frequency of Grade 5)	Worst severity: Grade 5	Extended
Subject Incidence by Special Criteria			
Events with an increase in the experimental arm of ≥5% (All Grades) or ≥2% (Grade 3/4)	SOC and PT (SOC per MedDRA standard, PT within SOC by decreasing difference between arms for All Grades)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
Non-serious adverse events of with an incidence of ≥ 5% in any arm (Any Grade)	SOC and PT (SOC per MedDRA standard, PT within SOC by decreasing difference between arms for All Grades)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
All	PT (descending frequency of difference in percent between the two arms for All Grades)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard
	PT (descending frequency of difference in percent between the two arms for Grade 3/4)	Worst severity: All Grades, Grade 3/4, Grade 4, Grade 5	Standard

In some of the above summaries classified by treatment exposure and relationship, the counts, and frequencies for the experimental arm (cabozantinib + atezolizumab) will be presented separately by individual treatment components as well as by combination.

The following data listings will also be provided with indicators for grade, relationship, seriousness, dose day of the event start/stop, days since last dose, actions taken with study treatment:

- All AEs
- Serious AEs
- All AEs that led to treatment discontinuation
- All AESI

8.2 Deaths

All subject deaths (Grade 5 TEAEs) will be summarized for all subjects in the Safety population.

Deaths will be summarized in 5 main categories as follows:

- All deaths
- Deaths within 30 days of last dose
- Deaths within 31 to 100 days of last dose
- Deaths within 100 days of last dose
- Deaths more than 100 days from last dose

In addition, under each category causality to study disease and relationship to study treatment will also be summarized.

All reported subject deaths will be listed.

8.3 Adverse Events of Special Interest

Adverse events of special interest (AESIs) consist of immune-mediated AEs associated with immune checkpoint inhibitor therapies, cases of potential drug induced liver injury (DILI) and suspected transmission of an infectious agent by the study treatment. Analysis of DILI is discussed in Section 8.4 and cases transmission of infectious agent by study treatment, if any, will be discussed in patient narratives.

Immune checkpoint inhibitor therapy is complicated by side effects known as immune-mediated adverse events classified as AESIs. The sponsor has defined a set of events to track these adverse events known to be associated with immune checkpoint inhibitors that have potentially serious consequences or were determined to warrant ongoing routine surveillance. Each AESI is a grouped clinical term comprising a set of pre-defined AE preferred terms for potentially immune-related events that are related pathophysiologically and provide a consistent, reproducible, and transparent compilation of safety information over time. The specific list of preferred terms employed for analyses will provided in the clinical study report.

The following summaries will be provided for AESIs for the P-Safety and Safety Populations. These summaries will be for all AESIs during the extended safety observation period.

TEAE included	Row-levels (sorted by)	Columns
Subject Incidence by SOC, Preferred Term and Severity		
All AESIs	AESI and PT (AESI, PT within AESI by descending frequency)	Worst severity: All Grades, Grade 3-4, Grade 4, Grade 5
AESIs leading to discontinuation of any treatment component	AESI and PT (AESI, PT within AESI by descending frequency)	Worst severity: Grade >=3, Grade 3, Grade 4, Grade 5
AESIs leading to discontinuation of all treatment components	AESI and PT (AESI, PT within AESI by descending frequency)	Worst severity: Grade >=3, Grade 3, Grade 4, Grade 5
AESIs leading to discontinuation of IV treatment components	AESI and PT (AESI, PT within AESI by descending frequency)	Worst severity: Grade >=3, Grade 3, Grade 4, Grade 5
AESIs leading to dose delay/reduction of any treatment component	AESI and PT (AESI, PT within AESI by descending frequency)	Worst severity: Grade >=3, Grade 3, Grade 4, Grade 5
AESIs leading to dose delay/reduction of all study treatment components	AESI and PT (AESI, PT within AESI by descending frequency)	Worst severity: Grade >=3, Grade 3, Grade 4, Grade 5
AESIs leading to dose delay/reduction of IV treatment component	AESI and PT (AESI, PT within AESI by descending frequency)	Worst severity: Grade >=3, Grade 3, Grade 4, Grade 5
AESIs leading to initiation of immune-modulating medication	AESI and PT (AESI, PT within AESI by descending frequency)	Worst severity: Grade >=3, Grade 3, Grade 4, Grade 5
AESIs leading to initiation of high dose of immune-modulating medication (defined in Section 6.4)	AESI and PT (AESI, PT within AESI by descending frequency)	Worst severity: Grade >=3, Grade 3, Grade 4, Grade 5
AESIs leading to initiation of immune-modulating medication and IV treatment discontinuation	AESI and PT (AESI, PT within AESI by descending frequency)	Worst severity: Grade >=3, Grade 3, Grade 4, Grade 5
All serious AESIs	AESI and PT (AESI , PT within AESI by descending frequency)	Worst severity: All Grades, Grade 3-4, Grade 4, Grade 5
Serious AESIs leading to initiation of immune-modulating medication	AESI and PT (AESI, PT within AESI by descending frequency)	Worst severity: Grade >=3, Grade 3, Grade 4, Grade 5

TEAE included	Row-levels (sorted by)	Columns
Serious AESIs leading to initiation of high dose of immune-modulating medication	AESI and PT (AESI, PT within AESI by descending frequency)	Worst severity: Grade >=3, Grade 3, Grade 4, Grade 5
Time to Event (per KM method only among those with event)		
AESIs	Time to first onset of AESIs by Group Term	
AESIs leading to initiation of immune-modulating medication	Time to first onset of AESIs by Group Term	
AESIs leading to initiation of high dose immune-modulating medication	Time to first onset of AESIs by Group Term	

8.4 Laboratory Assessments

8.4.1 Variables

The following treatment-emergent laboratory abnormalities will be summarized. A treatment emergent laboratory abnormality is defined as any laboratory abnormality with an onset date on or after the date of the first dose of study drug.

Category	Abnormality	SDTM LBTESTCD	Grading System
Hematology	WBC decreased	WBC	CTCAE
	ANC decreased	NEUT	CTCAE
	Lymphocytes increased Lymphocytes decreased	LYM	CTCAE
	Platelets decreased	PLAT	CTCAE
	Hemoglobin increased Hemoglobin decreased	HGB	CTCAE
Serum chemistry	Albumin decreased	ALB	CTCAE
	ALP increased	ALP	CTCAE
	Amylase increased	AMYLASE	CTCAE
	ALT increased	ALT	CTCAE
	AST increased	AST	CTCAE
	Calcium, corr increased Calcium, corr decreased	CACORR	CTCAE
	Creatinine increased	CREAT	CTCAE
	GGT increased	GGT	CTCAE

Category	Abnormality	SDTM LBTESTCD	Grading System
	Glucose increased Glucose decreased	GLUC	CTCAE
	LDH increased	LDH	Sponsor
	Lipase increased	LIPASE	CTCAE
	Magnesium increased Magnesium decreased	MG	CTCAE
	Phosphate decrease	PHOS	CTCAE
	Potassium increased Potassium decreased	K	CTCAE
	Sodium increased Sodium decreased	NA	CTCAE
	Total bilirubin increased	BILI	CTCAE
	Uric acid increased[2]	CYURIAC	CTCAE
Urine chemistry	UPCR increased	PROTCRT	Sponsor
Endocrinology[1]	Thyroid Stimulating Hormone increased Thyroid Stimulating Hormone decreased	TSH	HLN

[1] TSH is held in the SDTM "chemistry" laboratory category; will use HLN = high, low, normal classification based on normal range
[2] Uric acid increases will be graded only as Grade 1 or Grade 4. Grade 2 is not defined per CTCAE v4 and Grade 3 cannot be distinguished from Grade 1 based upon the result alone.

Sponsor-defined grades are to be applied to the following analytes:

LDH:

- Grade 1 if > ULN to ≤ 2 × ULN
- Grade 2 if > 2 × ULN to ≤ 3 × ULN
- Grade 3 if > 3 × ULN

UPCR:

- Grade 1 if ≥ 17.0 to ≤ 121.0 mg/mmol (≥ 0.15 to ≤ 1.0 mg/mg)
- Grade 2 if > 121.0 to ≤ 396.0 mg/mmol (>1.0 to < 3.5 mg/mg)
- Grade 3 if > 396.0 mg/mmol (> 3.5 mg/mg)

8.4.2 Analysis

All laboratory data parameters and visits will include flags for values above or below laboratory reference ranges. Toxicity grades will be assigned programmatically by applying the CTCAE v 5

guidelines. Only results with assessment dates through the end of the safety observation period (see Section 4.6) will be tabulated in summary tables.

Laboratory summaries will be presented in SI units. Continuous laboratory test results will be summarized by treatment group using descriptive statistics for actual values and for changes from baseline by scheduled visit. For test results which are below or above the quantification level, the imputed values as described in Appendix C will be used for deriving the grade and then summarized. Box-Whiskers plots may also be presented at each scheduled visit (with visits shown on x-axis) for some laboratory parameters for IDMC review purposes. Line graphs for mean change from baseline ± standard deviation will also be presented at each scheduled visit (with visits shown on x-axis) for the following laboratory parameters.

- Renal function
    - Blood urea nitrogen (BUN)
    - Creatinine
- Liver function panel
    - Total bilirubin
    - ALT
    - AST

Tables summarizing baseline laboratory values and the incidence of laboratory abnormalities by baseline and maximum post-baseline CTCAE grade over all records will be presented. In addition, the following summaries will also be presented:

A. Liver function abnormalities will be assessed as follows:
    - Shift from baseline based on normal ranges
    - Summaries of subjects meeting Hy's Law laboratory screening criteria as shown below:
        - $> 3 \times$ ULN (ALT or AST), $> 2 \times$ ULN Total Bilirubin, and $< 2 \times$ ULN ALP
        - $> 3 \times$ ULN (ALT or AST), $> 2 \times$ ULN Total Bilirubin, and $\geq 2 \times$ ULN ALP
        - $> 3 \times$ ULN (ALT or AST), $> 2 \times$ ULN Total Bilirubin

B. For renal failure surveillance, a summary of subjects meeting renal failure laboratory screening criteria as shown below will be provided:
    - Serum creatinine $\geq 3.0 \times$ ULN and $\geq 2.0 \times$ baseline value or
    - Estimated Glomerular Filtration Rate (eGFR) $\leq 50\%$ of the baseline value or
    - eGFR < 30 mL/min/1.73 m^2 and $\geq 25\%$ reduction from the baseline value

eGFR = 186 x (Creatinine in mmol per L / 88.4)$^{-1.154}$ x (Age)$^{-0.203}$ × (0.742 if female) × (1.210 if black) [from the UK CKD eGuide on the Renal Association website: http://egfrcalc.renal.org/]

C. Categorical summaries for baseline status for TSH and Free T4 will be presented. In addition, categorical summaries for post-baseline TSH and Free T4 status among subjects with TSH and Free T4 in the normal range at baseline will also be presented.

For descriptive summaries of change from baseline in continuous laboratory variables, only central lab results will be considered. For analyses of shift in grade from baseline or worst grade after baseline, all available results will be considered regardless of whether from the central or local lab.

Summary Type	Include central lab results?	Include local lab results?
Subject-Incidence of Treatment Emergent Laboratory Abnormalities in Selected Laboratory Tests by CTCAE Grade	Y	Y
Change from Baseline in Laboratory Values	Y	N
Shift from Baseline in Laboratory Values by CTCAE Grade	Y	Y
Worsening Shift of at Least 2 Grades from Baseline in Selected Laboratory Tests	Y	Y
Shift from Baseline in Laboratory Values by High/Low/Normal	Y	Y
Shift from Baseline in Laboratory Values by Sponsor-defined Grades	Y	Y
Laboratory Abnormalities with a subject-incidence of ≥5% (All Grades) or ≥2% (Grades 3-4) in either arm	Y	Y
Laboratory Abnormalities with a subject-incidence of ≥5% (All Grades) or ≥2% (Grades 3-4) in the experimental arm	Y	Y
Laboratory Abnormalities with an increase in subject-incidence of ≥5% (All Grades) or ≥2% (Grades 3-4) in the experimental arm	Y	Y

8.5 Vital Signs

8.5.1 Variables

The following vital signs will be summarized.

- Weight

- Systolic blood pressure (mmHg)
- Diastolic blood pressure (mmHg)

8.5.2 Analysis

Subject incidence of clinically meaningful changes since baseline for weight and blood pressure will be presented as shown below:

- Proportion of subjects with weight loss ≥ 10% after first dose
- Subjects with at least 2 post-baseline assessments and who got worse since baseline and met the following blood pressure criteria on 2 or more visits (need not be consecutive) after first dose (JNC criteria were modified to include single measurement per time point when triplicate assessments were unavailable; (Chobanian AV, 2003):

    - Pre-hypertension: SBP 120-139 mmHg or DBP 80-89 mmHg
    - Stage 1: SBP 140-159 mmHg or DBP 90-99 mmHg
    - Stage 2: (SBP ≥ 160 mmHg and DBP <120) or DBP 100-119 mmHg
    - Stage 3: DBP ≥ 120 mmHg

Only results with assessment dates through the end of the safety observation period (see Section 4.6) will be considered for the summaries.

8.6 ECOG Performance Status

The ECOG performance status of the subject will be assessed during screening and at each scheduled safety assessment starting on W1D1. For the purpose of evaluating safety, ECOG will be summarized as shift from baseline tables for, at minimum, the worst value recorded after the initiation of study treatment. Only results with assessment dates through the end of the safety observation period (see Section 4.6) will be tabulated in summary tables.

Frequencies of ECOG worsening of ≥ +1 and +2 change from baseline to worst value after first dose will also be summarized.

8.7 Electrocardiogram (ECG)

Only results with assessment dates through the end of the safety observation period (see Section 4.6) will be considered for summaries. The following categorical summaries will be presented per investigator:

- Number of subjects with triplicate average QTc > 500 ms or increase in triplicate average QTc from baseline of > 60 ms after first dose per investigator
    - number of subjects with triplicate average QTc > 500 ms after first dose per investigator
        - number of cases referred for independent review (ERT)
        - number of cases confirmed as QTc > 500 ms by ERT.
    - number of subjects with increase in triplicate average QTc from baseline of > 60 ms after first dose per investigator
- number of subjects with increase in triplicate average QTc from baseline of > 30 ms after first dose per investigator

For the above summaries, the most-recent average value from triplicate measurements taken before first dose will be used as baseline. If no triplicate measurements were taken before first dose, the most-recent single or duplicate value taken before first dose will be used as baseline. If > 3 measurements are taken for an assessment, all will be included in the average.

8.8 Impact of COVID-19 Pandemic

Summaries and analyses to describe and/or assess the impact of the COVID-19 pandemic will be included in a future amendment to this plan and/or provided in the study report. These may include tabulations of COVID-19 related protocol deviations and patterns of missing data, summaries of COVID-19 AEs, the addition of a per-protocol population, and comparative analyses of selected endpoint before, during and after the pandemic.

9 IDENTIFICATION AND SUMMARY OF PROTOCOL DEVIATIONS

Important protocol deviations as pre-specified in the study Protocol Deviation Management Plan. In accordance with ICH E3, Important eligibility deviations (as documented on study CRFs and Important post-randomization protocol deviations tracked in the study clinical trial management system (CTMS) will be identified and listed separately by study center and subject. Important deviations will be summarized for the PITT and ITT populations by deviation code (a standardized description e.g. “did not satisfy eligibility criteria” or “received prohibited medication”)

10 DATA QUALITY ASSURANCE

The Clinical, Data Management, Biostatistics, and Medical Writing departments at Exelixis and designees will work diligently and collaboratively to ensure that the data collected and analyzed for this study are of the highest quality. Controls including Case Report Form (CRF) design, CRF annotation, database design, data-entry processes, data validation programs, discrepancy management, consistent use of medical coding, data extraction processes, and database locking processes ensure the quality of the data. These controls are assessed at regular intervals during a trial.

11 REFERENCES

Chobanian AV, B. G. (2003, May). The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure: The JNC 7 Report. *JAMA, 289*(19), 2560-72.

Clopper CJ, P. E. (1934). The use of Confidence or Fiducial Limits Illustrated in the Case of Binomial. *Biometrika, 26*(4), 404-413.

Common Terminology Criteria for Adverse Events, version 5.0. (2017, November 27). USDHHS, NIH, NCI.

Eisenhauer, E. T. (2009). New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *European Journal of Cancer, 45*, 228-247.

FDA Guidance for Industry: Multiple Endpoints in Clinical Trials (draft). (2017, April).

Hessel C, M. M. (2016). Evaluation of the novel "trial within a trial" design of METEOR, a randomized phase 3 trial of cabozantinib versus everolimus in patients with advanced RCC. *ESMO.*

ICH Harmonized Guideline E9 (R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials (Draft). (2017). ICH.

ICH Harmonized Guideline E9 Statistical Principles for Clinical Trials. (1998). *ICH.*

Lencioni, R. a. (2010). Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. *Semin Liver Dis, 30*(1), 52-60.

Reenen MV, J. B. (2019). EQ-5D-5L User Guide. *3.0*.

Appendix A: Date Imputation Rules

Incomplete Cancer Diagnosis Date

If *year* is missing (or completely missing): do not impute

If only *day* is missing: set to 15th of the month.

If *day* and *month* are missing: set to July 1st.

If either imputation rule above results in a diagnosis date > informed consent:

set diagnosis date to the date of informed consent - 1.

Incomplete Adverse Event Onset Date

Assumption: For on-study Adverse Events.

If *year* is missing (or completely missing): set to the date of first dose.

If (*year* is present and *month* and *day* are missing) or (*year* and *day* are present and *month* is missing):

If *year* = year of first dose: set the date to the first dose date.

If *year* < year of first dose: set *month* and *day* to December 31st.

If *year* > year of first dose: set *month* and *day* to January 1st.

If *month* and *year* are present and *day* is missing:

If *year* = year of first dose, and:

If *month* = month of first dose: set *day* to day of first dose.

If *month* < month of first dose: set *day* to last day of *month*.

If *month* > month of first dose: set *day* to 1st day of *month*.

If *year* < year of first dose: set *day* to last day of month.

If *year* > year of first dose: set *day* to 1st day of month.

For all other cases: set to date of first dose.

Incomplete Concomitant Medication Start Date

If *year* is missing (or completely missing): do not impute.

If (*year* is present and *month* and *day* are missing) or (*year* and *day* are present and *month* is missing):

Set *month* and day to January 1st.

If *year* and *month* are present and *day* is missing:

Set *day* to 1st day of month.

Incomplete Concomitant Medication End Date

If *year* is missing (or completely missing): do not impute.

If (*year* is present and *month* and *day* are missing) or (*year* and *day* are present and *month* is missing):

Set *month* and day to December 31st.

If *year* and *month* are present and *day* is missing:

Set *day* to last day of the month.

Incomplete Subsequent Anticancer Therapy Start Date

Assumption: Anticancer therapies reported on the Subsequent Anticancer Therapy CRF.

If *year* is missing (or completely missing): set to date decision to discontinue study treatment + 1

If (*year* is present and *month* and *day* are missing) or (*year* and *day* are present and *month* is missing):

If *year* > year of date of decision to discontinue study treatment: Set *month* and *day* to January 1st.

If *year* = year of date of decision to discontinue study treatment: Set *month* and *day* to date of decision to discontinue study treatment + 1

If *year* and *month* are present and *day* is missing:

Set *day* to 1st day of month if the resulting imputed date is greater than date of decision to discontinue study treatment or if the month is before the month of date of decision to discontinue study treatment and year is same or before the year of the date of decision to discontinue study treatment. Otherwise set the imputed date to date of decision to discontinue study treatment + 1

Incomplete Death Date

Identify date of last known alive (LA) prior to death from the following:

1. Date of decision to discontinue study treatment from End of Treatment CRF
2. Date of last radiographic assessment from End of Radiographic Follow Up CRF
3. Date last known alive from Survival Follow Up CRF
4. Date of last lab assessment from the Labs dataset

If *year* is missing (or completely missing): set to date of LA + 1

If only *day* is missing: set to the maximum of the first of month or LA + 1

If *month* and *day* are missing:

If *year* of LA = year of death

Set death date to date of LA + 1

If *year* of most-recent contact < year of death

Set *month* and *day* to Jan 1st.

Incomplete Study Treatment Start Date

Define previous sequential dosing "milestone" as the latest of previous dose stop date, previous dose hold stop date, date of first dose or randomization date.

If *year* is missing (or completely missing): set to date of previous sequential dosing "milestone" + 1

If (*year* is present and *month* and *day* are missing) or (*year* and *day* are present and *month* is missing): set to January 1st

If *year* and *month* are present and *day* is missing: set to the first day of the month

If the imputed date is before the previous sequential dosing "milestone": set to the date of previous sequential dosing "milestone" + 1

Incomplete Study Treatment Stop Date

Define next sequential dosing "milestone" as the earliest of next dose start date, next dose hold start date, date of last dose from EOT CRF or the cutoff date.

If *year* is missing (or completely missing): set to date of next sequential dosing "milestone" - 1

If (*year* is present and *month* and *day* are missing) or (*year* and *day* are present and *month* is missing): set to December 31st

If *year* and *month* are present and *day* is missing: set to the last day of the month

If the imputed date is after the next sequential dosing "milestone": set to the date of next sequential dosing "milestone" - 1

Appendix B: Rounding Rules for Reported Percentages

For percentages ≥ 10%:

- Values ≥ X.5 or above round to X+1.
- Values >X but <X.5 round to X.

For percentages <10%:

- Values ≥ X.Y5 or above round to X.Y+0.1.
- Values >X.Y but <X.Y5 round to X.Y.

Appendix C: Imputation Rules for Laboratory Values Outside of Quantification Range

- Lab values below the lower level of quantification (LLQ) that are reported as "< LLQ" or "≤ LLQ" in the database will be imputed by LLQ × 0.99 for analysis purposes. However, the original value will also be maintained.

- Lab values above the upper level of quantification (ULQ) that are reported as "> ULQ" or "≥ ULQ" in the database will be imputed by ULQ × 1.01 for analysis purposes. However, the original value will also be maintained.

Appendix D: EQ-5D-5L Index Value Conversion Guidelines

The EQ-index conversion algorithm (EQ-5D-5L User Guide 2.1, April 2015. Available from: http://www.euroqol.org/about-eq-5d/publications/user-guide.html):

- Calculate ***health state***
  - Each of the 5 dimensions comprising the EQ-5D descriptive system is divided into 5 levels of perceived problems:
    - Level 1: indicating no problem
    - Level 2: indicating slight problems
    - Level 3: indicating moderate problems
    - Level 4: indicating severe problems
    - Level 5: indicating extreme problems
  - A unique health state is defined by combining 1 level from each of the 5 dimensions.
    For example, state 11111 indicates no problems on any of the 5 dimensions, while state 12345 indicates no problems with mobility, slight problems with washing or dressing, moderate problems with doing usual activities, severe pain or discomfort and extreme anxiety or depression. Note that missing values will be coded as '9'. Ambiguous values will be treated as missing values.
- ***EQ-index values*** for each country = ***health state * the country specific conversion factors*** for each dimension (EQ-5D-5L Index Value Calculator, version 1)

Appendix E: Estimands Terminology

Source: ICH E9 R1

Estimand Attributes:

#	Estimand Attribute
1	The population, that is, the patients targeted by the scientific question.
2	The variable (or endpoint), to be obtained for each patient, that is required to address the scientific question.
3	The specification of how to account for intercurrent events to reflect the scientific question of interest.
4	The population-level summary for the variable which provides, as required, a basis for a comparison between treatment conditions.

Strategies for Addressing Intercurrent Event(s):

Strategy	Description
Treatment policy	The occurrence of the intercurrent event is irrelevant: the value for the variable of interest is used regardless of whether or not the intercurrent event occurs.
Composite	The occurrence of the intercurrent event is taken to be a component of the variable, i.e. the intercurrent event is integrated with one or more other measures of clinical outcome as the variable of interest.
Hypothetical	A scenario is envisaged in which the intercurrent event would not occur: the value to reflect that scientific question of interest is that which the variable would have taken in the hypothetical scenario defined.
Principal stratum	The target population might be taken to be the principal stratum (see Glossary) in which an intercurrent event would not occur.
While on treatment	Response to treatment prior to the occurrence of the intercurrent event is of interest.

Certificate Of Completion

Envelope Id: 922FC5DAA3AD44808433C27934A373A2
Status: Completed
Subject: Please DocuSign: X184-312 SAP Version 1.0 Final .docx
Source Envelope:
Document Pages: 83
Signatures: 5
Envelope Originator:
Certificate Pages: 6
Initials: 0
PPD
AutoNav: Enabled
PPD
EnvelopeId Stamping: Disabled
IP Address: PPD
Time Zone: (UTC-08:00) Pacific Time (US & Canada)

Record Tracking

Status: Original
05-Feb-2021 | 11:11
Holder: PPD
Location: DocuSign

Signer Events	Signature	Timestamp
PPD Exelixis, Inc. - Part 11 Security Level: Email, Account Authentication (Required)	PPD Signature Adoption: Pre-selected Style Signature ID: F6D62EDE-7F39-42B4-865A-0FD29DA3F21A Using IP Address: PPD With Signing Authentication via DocuSign password With Signing Reasons (on each tab): I approve this document	Sent: 05-Feb-2021 | 11:19 Viewed: 05-Feb-2021 | 11:45 Signed: 05-Feb-2021 | 11:46
Electronic Record and Signature Disclosure: Accepted: 27-Jul-2020 | 13:19 ID: fceba3af-47d2-4916-92ff-d57ec274972c		
PPD Exelixis, Inc. Security Level: Email, Account Authentication (Required)	PPD Signature Adoption: Pre-selected Style Signature ID: 8CC40794-5B10-48C8-BA7C-95F84AA1FB03 Using IP Address: PPD With Signing Authentication via DocuSign password With Signing Reasons (on each tab): I approve this document	Sent: 05-Feb-2021 | 11:19 Viewed: 05-Feb-2021 | 11:58 Signed: 05-Feb-2021 | 11:58
Electronic Record and Signature Disclosure: Accepted: 14-Aug-2020 | 18:19 ID: b211a1af-2d3e-4856-ae6d-1d13a22a43d1		

Signer Events	Signature	Timestamp
PPD Exelixis, Inc. - Part 11 Security Level: Email, Account Authentication (Required) **Electronic Record and Signature Disclosure:** Accepted: 31-Jul-2020 | 08:34 ID: 53a02783-ac47-4a94-b936-d7510da867c9	PPD Signature Adoption: Pre-selected Style Signature ID: 89F2FB70-CDBF-4FFF-9A9F-E6C1DE23844E Using IP Address: PPD With Signing Authentication via DocuSign password With Signing Reasons (on each tab): I approve this document	Sent: 05-Feb-2021 | 11:19 Viewed: 05-Feb-2021 | 11:20 Signed: 05-Feb-2021 | 11:20
PPD Exelixis, Inc. - Part 11 Security Level: Email, Account Authentication (Required) **Electronic Record and Signature Disclosure:** Accepted: 17-Aug-2020 | 16:04 ID: 49dea567-2b8a-404e-ab09-a33f6804f7a5	PPD Signature Adoption: Pre-selected Style Signature ID: F73C78B2-68C4-42D0-A21F-65C2B8C2BE8A Using IP Address: PPD With Signing Authentication via DocuSign password With Signing Reasons (on each tab): I approve this document	Sent: 05-Feb-2021 | 11:19 Viewed: 05-Feb-2021 | 14:16 Signed: 05-Feb-2021 | 14:16
PPD Exelixis, Inc. Security Level: Email, Account Authentication (Required) **Electronic Record and Signature Disclosure:** Accepted: 28-Jul-2020 | 14:59 ID: ecc165c2-4dae-4f87-8c5f-33e0319cced7	PPD Signature Adoption: Uploaded Signature Image Signature ID: 43644AE1-51D2-4DAE-81FA-F4BB29A1CCC9 Using IP Address: PPD With Signing Authentication via DocuSign password With Signing Reasons (on each tab): I approve this document	Sent: 05-Feb-2021 | 11:19 Viewed: 05-Feb-2021 | 11:44 Signed: 05-Feb-2021 | 11:44

In Person Signer Events	Signature	Timestamp

Editor Delivery Events	Status	Timestamp

Agent Delivery Events	Status	Timestamp

Intermediary Delivery Events	Status	Timestamp

Certified Delivery Events	Status	Timestamp

Carbon Copy Events	Status	Timestamp

Witness Events	Signature	Timestamp

Notary Events	Signature	Timestamp

Envelope Summary Events	Status	Timestamps
Envelope Sent	Hashed/Encrypted	05-Feb-2021 | 11:19
Certified Delivered	Security Checked	05-Feb-2021 | 11:44
Signing Complete	Security Checked	05-Feb-2021 | 11:44
Completed	Security Checked	05-Feb-2021 | 14:16

Payment Events	Status	Timestamps

Electronic Record and Signature Disclosure

Note: Electronic signatures executed in DocuSign are the legally binding equivalent of your traditional hand written signatures and subject to the electronic requirements of 21 CFR Part 11 and Annex 11.

ELECTRONIC RECORD AND SIGNATURE DISCLOSURE

From time to time, Exelixis, Inc. - Part 11 (we, us or Company) may be required by law to provide to you certain written notices or disclosures. Described below are the terms and conditions for providing to you such notices and disclosures electronically through the DocuSign system. Please read the information below carefully and thoroughly, and if you can access this information electronically to your satisfaction and agree to this Electronic Record and Signature Disclosure (ERSD), please confirm your agreement by selecting the check-box next to 'I agree to use electronic records and signatures' before clicking 'CONTINUE' within the DocuSign system.

Getting paper copies

At any time, you may request from us a paper copy of any record provided or made available electronically to you by us. You will have the ability to download and print documents we send to you through the DocuSign system during and immediately after the signing session and, if you elect to create a DocuSign account, you may access the documents for a limited period of time (usually 30 days) after such documents are first sent to you. After such time, if you wish for us to send you paper copies of any such documents from our office to you, you will be charged a $0.00 per-page fee. You may request delivery of such paper copies from us by following the procedure described below.

Withdrawing your consent

If you decide to receive notices and disclosures from us electronically, you may at any time change your mind and tell us that thereafter you want to receive required notices and disclosures only in paper format. How you must inform us of your decision to receive future notices and disclosure in paper format and withdraw your consent to receive notices and disclosures electronically is described below.

Consequences of changing your mind

If you elect to receive required notices and disclosures only in paper format, it will slow the speed at which we can complete certain steps in transactions with you and delivering services to you because we will need first to send the required notices or disclosures to you in paper format, and then wait until we receive back from you your acknowledgment of your receipt of such paper notices or disclosures. Further, you will no longer be able to use the DocuSign system to

receive required notices and consents electronically from us or to sign electronically documents from us.

All notices and disclosures will be sent to you electronically

Unless you tell us otherwise in accordance with the procedures described herein, we will provide electronically to you through the DocuSign system all required notices, disclosures, authorizations, acknowledgements, and other documents that are required to be provided or made available to you during the course of our relationship with you. To reduce the chance of you inadvertently not receiving any notice or disclosure, we prefer to provide all of the required notices and disclosures to you by the same method and to the same address that you have given us. Thus, you can receive all the disclosures and notices electronically or in paper format through the paper mail delivery system. If you do not agree with this process, please let us know as described below. Please also see the paragraph immediately above that describes the consequences of your electing not to receive delivery of the notices and disclosures electronically from us.

How to contact Exelixis, Inc. - Part 11:

You may contact us to let us know of your changes as to how we may contact you electronically, to request paper copies of certain information from us, and to withdraw your prior consent to receive notices and disclosures electronically as follows:
To contact us by email send messages to: eoleary@exelixis.com

To advise Exelixis, Inc. - Part 11 of your new email address

To let us know of a change in your email address where we should send notices and disclosures electronically to you, you must send an email message to us at eoleary@exelixis.com and in the body of such request you must state: your previous email address, your new email address. We do not require any other information from you to change your email address.

If you created a DocuSign account, you may update it with your new email address through your account preferences.

To request paper copies from Exelixis, Inc. - Part 11

To request delivery from us of paper copies of the notices and disclosures previously provided by us to you electronically, you must send us an email to eoleary@exelixis.com and in the body of such request you must state your email address, full name, mailing address, and telephone number. We will bill you for any fees at that time, if any.

To withdraw your consent with Exelixis, Inc. - Part 11

To inform us that you no longer wish to receive future notices and disclosures in electronic format you may:

i. decline to sign a document from within your signing session, and on the subsequent page, select the check-box indicating you wish to withdraw your consent, or you may;

ii. send us an email to eoleary@exelixis.com and in the body of such request you must state your email, full name, mailing address, and telephone number. We do not need any other information from you to withdraw consent.. The consequences of your withdrawing consent for online documents will be that transactions may take a longer time to process..

Required hardware and software

The minimum system requirements for using the DocuSign system may change over time. The current system requirements are found here: https://support.docusign.com/guides/signer-guide-signing-system-requirements.

Acknowledging your access and consent to receive and sign documents electronically

To confirm to us that you can access this information electronically, which will be similar to other electronic notices and disclosures that we will provide to you, please confirm that you have read this ERSD, and (i) that you are able to print on paper or electronically save this ERSD for your future reference and access; or (ii) that you are able to email this ERSD to an email address where you will be able to print on paper or save it for your future reference and access. Further, if you consent to receiving notices and disclosures exclusively in electronic format as described herein, then select the check-box next to 'I agree to use electronic records and signatures' before clicking 'CONTINUE' within the DocuSign system.

By selecting the check-box next to 'I agree to use electronic records and signatures', you confirm that:

- You can access and read this Electronic Record and Signature Disclosure; and
- You can print on paper this Electronic Record and Signature Disclosure, or save or send this Electronic Record and Disclosure to a location where you can print it, for future reference and access; and
- Until or unless you notify Exelixis, Inc. - Part 11 as described above, you consent to receive exclusively through electronic means all notices, disclosures, authorizations, acknowledgements, and other documents that are required to be provided or made available to you by Exelixis, Inc. - Part 11 during the course of your relationship with Exelixis, Inc. - Part 11.

X184-312: Addendum to Statistical Analysis Plan Version 1.0

A Randomized, Controlled Phase 3 Study of Cabozantinib (XL184) in Combination with Atezolizumab versus Sorafenib in Subjects with Advanced Hepatocellular Carcinoma Who Have Not Received Previous Systemic Anticancer Therapy

SAP Finalization Date: February 5, 2021
Addendum Date: June 3, 2021

Prepared By	PPD Exelixis
Approved By	PPD Exelixis

Approved By	PPD  Exelixis
Approved By	PPD  PPD , Exelixis
Approved By	PPD  PPD , Exelixis

1 STATISTICAL ANALYSIS PLAN ADDENDUM (GLOBAL)

1.1 Rationale for Addendum

This addendum provides clarity about the pre-specified inferential analysis for event-driven time-to-event endpoints when the following conditions arise due to circumstances such as faster than expected accrual, shorter than expected event times, and/or event ascertainment lag:

- At an interim analysis, the actual information fraction (IF) exceeds the planned IF
- At a final analysis, the total events exceed the planned number of events

2 CONVENTION

In both circumstances:

- The total information/events available at time of the analysis is prespecified as the primary analysis for inference testing and alpha-spending
- The planned information/planned number of events described in the Statistical Analysis Plan will be summarized descriptively as a supportive analysis.