Official Title: A Phase III, Open-Label, Randomized Study of Atezolizumab in Combination With Bevacizumab Compared With Sorafenib in Patients With Untreated Locally Advanced or Metastatic Hepatocellular Carcinoma

NCT Number: NCT03434379

Document Date: SAP Version 1: 22-February-2019

STATISTICAL ANALYSIS PLAN

TITLE: **A PHASE III, OPEN-LABEL, RANDOMIZED STUDY OF ATEZOLIZUMAB IN COMBINATION WITH BEVACIZUMAB COMPARED WITH SORAFENIB IN PATIENTS WITH UNTREATED LOCALLY ADVANCED OR METASTATIC HEPATOCELLULAR CARCINOMA**

PROTOCOL NUMBER: YO40245

STUDY DRUG: Atezolizumab
Bevacizumab

VERSION NUMBER: 1

IND NUMBER: 135913

EUDRACT NUMBER: 2017-003691-31

SPONSOR: F. Hoffmann-La Roche Ltd.

PLAN PREPARED BY: [redacted] Ph.D.

DATE FINAL: See electronic date stamp below

STATISTICAL ANALYSIS PLAN APPROVAL

Name	Reason for Signing	Date and Time (UTC)
[redacted]	Company Signatory	22-Feb-2019 16:19:43

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

1. BACKGROUND

This Statistical Analysis Plan (SAP) is based on Protocol YO40245 (IMbrave150), "A Phase III, randomized, multicenter, open-label, two-arm study designed to evaluate the efficacy and safety of atezolizumab + bevacizumab versus sorafenib in patients with untreated locally advanced or metastatic hepatocellular carcinoma (HCC)". This SAP provides details of the planned analyses and statistical methods for Study YO40245 to support potential registration for atezolizumab in combination with bevacizumab for patients with unresectable or advanced HCC who have received no prior systemic therapy. The background for the study can be found in the study protocol. Analyses that are beyond those outlined in the protocol are delineated in this document.

In this study, the **global population** will include all patients enrolled during the global enrollment phase, and the **China subpopulation** will include all patients enrolled at sites that are recognized by the China Food and Drug Administration during both the global enrollment phase and the extended China enrollment phase. This SAP documents details of the statistical analyses planned for the global study population (Section 4) and the China subpopulation (Section 5).

2. STUDY DESIGN

This is a Phase III, randomized, multicenter, open-label, two-arm study designed to evaluate the efficacy and safety of atezolizumab + bevacizumab versus sorafenib in patients with locally advanced or metastatic HCC who have received no prior systemic treatment.

This study will randomize approximately 480 eligible patients in a 2:1 ratio to one of the two treatment regimens as shown in Table 1.

Table 1 Study YO40245 Treatment Arms

Treatment Arm	Dose, Route, and Regimen [a]
A	Atezolizumab 1200 mg IV infusions Q3W (dosed in 3-week cycles) + bevacizumab 15 mg/kg Q3W (dosed in 3-week cycles)
B	Sorafenib 400 mg PO, BID, continuously

BID = twice per day; IV = intravenous; PO = by mouth; Q3W = every 3 weeks.

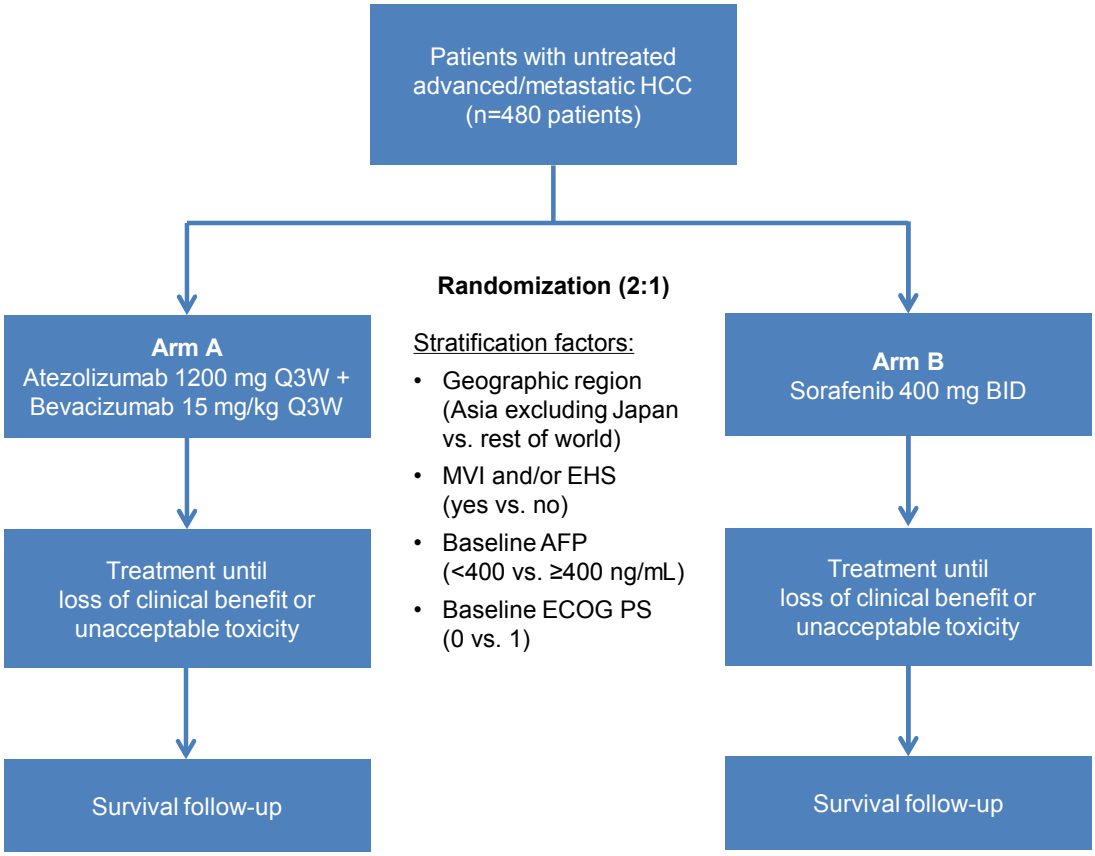
Note: No dose modification for atezolizumab or bevacizumab is allowed.

[a] 21-day treatment cycles until unacceptable toxicity or loss of clinical benefit.

Randomization details, including stratification factors, are outlined in Section 3.1.

Figure 1 illustrates the study design.

Figure 1 Study Schema



AFP = α-fetoprotein; BID = twice per day; ECOG PS = Eastern Cooperative Oncology Group Performance Status; EHS = extrahepatic spread; HCC = hepatocellular carcinoma; MVI = macrovascular invasion; PS = Performance Status; Q3W = every 3 weeks.

Patients will receive atezolizumab and/or bevacizumab or sorafenib until unacceptable toxicity or loss of clinical benefit as determined by the investigator after an integrated assessment of radiographic and biochemical data, and clinical status (e.g., symptomatic deterioration such as pain secondary to disease). In the absence of unacceptable toxicity, patients who meet criteria for disease progression per Response Evaluation Criteria in Solid Tumors (RECIST) v1.1 while receiving atezolizumab and/or bevacizumab or sorafenib will be permitted to continue the study treatment if they meet all of the following criteria:

- Evidence of clinical benefit, as determined by the investigator following a review of all available data
- Absence of symptoms and signs (including laboratory values, such as new or worsening hypercalcemia) indicating unequivocal progression of disease
- Absence of decline in Eastern Co-operative Oncology Group Performance (ECOG) Performance Status that can be attributed to disease progression

- Absence of tumor progression at critical anatomical sites (e.g., leptomeningeal disease) that cannot be managed by protocol-allowed medical interventions

The co-primary efficacy endpoints of this study are overall survival (OS) and progression-free survival as assessed by Independent Review Facility (PFS-IRF) according to RECIST v1.1. See Section 4.4.1 for further details on the co-primary efficacy endpoints, as well as other subsections of 4.4 on the secondary efficacy endpoints and the safety, pharmacokinetic (PK), immunogenicity, and exploratory endpoints. Unless otherwise specified, the term "PFS" in this SAP refers to PFS-IRF according to RECIST v1.1.

2.1 PROTOCOL SYNOPSIS

The protocol synopsis can be found in Appendix 1, including the study objectives, inclusion and exclusion criteria, outcome measures, and statistical methods as stated in the protocol. Schedule of Activities is in Appendix 2 and schedule of biomarker, pharmacokinetic, and immunogenicity samples is in Appendix 3.

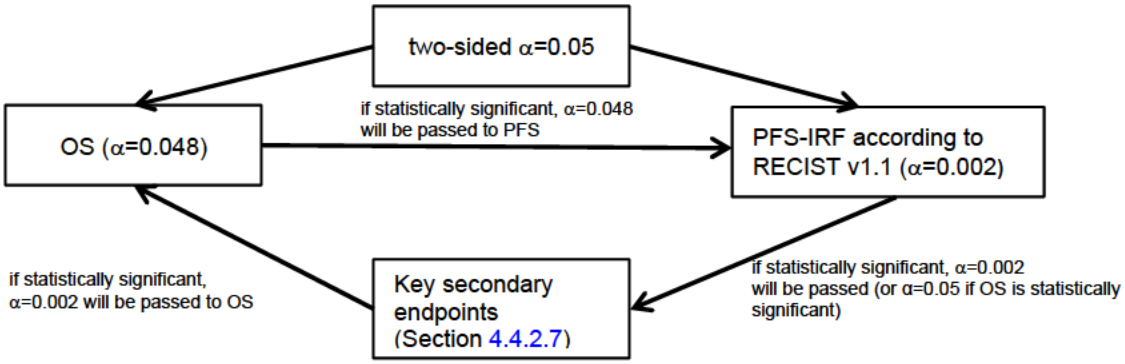
2.2 DETERMINATION OF SAMPLE SIZE

Approximately 480 patients are planned for enrollment during the global enrollment phase.

2.2.1 Type I Error Control

The overall type I error rate for this study will be strongly controlled at a two-sided significance level of 0.05 by a graphical approach, i.e., alpha splitting and recycling (Bretz et al. 2009, Burman et al. 2009). The overall two-sided significance level of 0.05 will be split into a two-sided significance level of 0.048 for the testing of OS and a two-sided significance level of 0.002 for the testing of PFS initially. If OS is statistically significant, the allocated two-sided significance level of 0.048 can be recycled to PFS such that PFS can be tested at a two-sided significance level of 0.05 instead of 0.002. If the analysis of PFS is statistically significant, then the two-sided significance level of 0.002 (or 0.05 if OS is statistically significant) will be recycled to key secondary endpoints to be formally tested in a hierarchical fashion (See Section 4.4.2.7 for details on secondary endpoints to be included for formal statistical testing). If PFS and all key secondary endpoints are statistically significant at a two-sided significance level of 0.002, then OS can be tested at a two-sided significance level of 0.05 instead of 0.048; the detailed testing boundaries are provided in Section 2.3. An overview of the type I error rate control strategy for the co-primary and key secondary efficacy endpoints is shown in Figure 2.

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



OS = overall survival; PFS = progression-free survival; PFS-IRF = progression-free survival as assessed by Independent Review Facility; RECIST = Response Evaluation Criteria in Solid Tumors.

2.2.2 Co-Primary Endpoint: OS

The sample size of the study was determined based on the number of deaths required from the patients randomized in the global enrollment phase to demonstrate efficacy in terms of OS. To detect an improvement in OS using a log-rank test at a two-sided significance level of 0.048, approximately 312 deaths will be required at the final OS analysis to achieve an overall 80% power assuming a target hazard ratio (HR) of 0.71 (median OS improvement vs. control of 4.9 months). The minimum detectable difference (MDD) of OS is an HR of 0.783 (median OS improvement vs. control of 3.3 months). This analysis is expected to occur approximately 33 months after first-patient in (FPI).

The calculation of sample size and estimates of the OS analysis timelines are based on the following assumptions for the intent-to-treat (ITT) population:

- Patients will be randomized to the atezolizumab + bevacizumab and sorafenib arms in a 2:1 ratio
- OS follows a one-piece exponential distribution
- The median OS in the control arm is 12 months
- The stopping boundaries of two interim analyses and the final analysis of OS use the O'Brien-Fleming boundaries approximated using the Lan-DeMets method (DeMets and Lan 1994)
- The dropout rate is 5% for the atezolizumab + bevacizumab arm and 10% for the sorafenib arm over 12 months for OS
- The recruitment of approximately 480 patients will take place over approximately 10 months

2.2.3 Co-Primary Endpoint: PFS by IRF-Assessment (PFS-IRF) per RECIST v1.1

To detect an improvement in PFS-IRF per RECIST v1.1 using a log-rank test at a two-sided significance level of 0.002, approximately 308 events will be required for the primary PFS analysis, which will provide approximately 97% power with a target HR of 0.55 (median PFS improvement vs. control of 3.3 months). The MDD is a PFS HR of 0.688 (median PFS improvement vs. control of 1.8 months). The clinical cutoff date for this primary PFS analysis is expected to occur approximately 16 months after the first patient was enrolled in the study. The following assumptions were made for PFS based on the ITT population:

- Patients will be randomized to the atezolizumab + bevacizumab and sorafenib arms in a 2:1 ratio
- PFS follows a one-piece exponential distribution
- The median PFS in the control arm is 4 months
- The dropout rate is 5% for the atezolizumab plus bevacizumab arm and 10% for the sorafenib arm over 12 months for PFS
- The recruitment of approximately 480 patients will take place over approximately 10 months.

See Section 5 for sample size considerations for the China subpopulation.

2.3 INTERIM AND PRIMARY ANALYSIS TIMING

2.3.1 Primary Analysis Timing

The primary analysis of the co-primary endpoint PFS-IRF per RECIST v1.1 based on randomized patients enrolled during the global enrollment phase will be performed when approximately 308 PFS events are observed. This is to have a minimal follow-up of approximately 6 months for patients randomized in the global enrollment phase (i.e., to have a clinical cutoff date at approximately 6 months after global last patient in [LPI]).

2.3.2 Interim Analyses Timing

No interim analysis (IA) is planned for the co-primary endpoint PFS in this study.

Two interim analyses of OS will be performed. The first IA will be performed at the time of the primary PFS analysis, estimated to occur at approximately 16 months after FPI. It is anticipated that at this time, approximately 172 deaths will have been observed. The respective MDD OS HR is 0.633 (median OS improvement vs. control of 6.9 months). The second OS IA is planned to be conducted when approximately 243 deaths are accumulated, estimated to occur at approximately 24 months after FPI. The respective MDD OS HR is 0.728 (median OS improvement vs. control of 4.6 months).

To control the type I error for OS, the stopping boundaries for the OS interim and final analyses are to be computed with use of the Lan–DeMets approximation to the

O'Brien-Fleming boundary (DeMets and Lan 1994) as shown in Table 2. The projected p-value and MDD boundaries were computed based on the number of OS events required at each OS analysis; the actual boundaries will be calculated at the time of OS analysis based on the observed information fraction, i.e., actual number of events observed at time of analysis over the total planned target number of events in the ITT population.

The planned IAs for OS will be conducted by the Sponsor.

Table 2 Analysis Timing and Stopping Boundaries for Overall Survival

Analysis Timing	Planned Information Fraction	Number of Events/ Analysis timing (estimated)	Stopping Boundary (Two-Sided p-Value)	
			Alpha can be recycled to OS (i.e. OS alpha = 0.05)	Alpha cannot be recycled to OS (i.e. OS alpha = 0.048)
1st OS interim analysis	55%	172/16 months*	MDD HR ≤ 0.636 (p-value ≤ 0.005)	MDD HR ≤ 0.633 (p-value ≤ 0.005)
2nd OS interim analysis	78%	243/24 months	MDD HR ≤ 0.73 (p-value ≤ 0.021)	MDD HR ≤ 0.728 (p-value ≤ 0.02)
OS final analysis	100%	312/33 months	MDD HR ≤ 0.784 (p-value ≤ 0.043)	MDD HR ≤ 0.783 (p-value ≤ 0.041)

HR = hazard ratio; MDD = minimum detectable difference; OS = overall survival; PFS = progression-free survival.

Analysis timing shown in the table is projected based on protocol assumptions. Actual timing depends on the exact time that the required events have accrued.

*The 1st OS interim analysis will be conducted when approximately 308 PFS events have occurred. It is anticipated that approximately 172 OS events have been observed at time of primary PFS analysis.

MDD HR is estimated based on proportional hazard assumption.

3. STUDY CONDUCT

3.1 RANDOMIZATION

Eligible patients will be randomized in a 2:1 ratio to receive either atezolizumab + bevacizumab or sorafenib with the use of a stratified permuted-block randomization. The randomization will be stratified according to the following stratification factors:

- Geographic region (Asia excluding Japan vs. rest of world)
- Macrovascular invasion and/or extrahepatic spread (presence vs. absence)
- Baseline AFP (α-fetoprotein [< 400 vs. ≥ 400 ng/mL])
- ECOG Performance Status (0 vs. 1)

3.2 DATA MONITORING

An independent Data Monitoring Committee (iDMC) will be implemented to evaluate safety during the study on a regular basis. All summaries and analyses by treatment arm for the iDMC review are prepared by an external independent Data Coordinating Center. Members of the iDMC are external to the Sponsor and follow a separate iDMC Charter that outlines their roles and responsibilities, as well as a detailed monitoring plan. The meetings are scheduled approximately every 6 months for the review of the accrued safety data, with the first review meeting based on a data cut approximately 6 months after FPI. iDMC membership will extend until the study is unblinded to the Sponsor. Further details will be provided in the iDMC charter.

All efficacy analyses including primary PFS analysis, interim and final OS analyses will be conducted by the Sponsor.

3.3 INDEPENDENT REVIEW FACILITY

An IRF will be used to conduct a blinded centralized radiology review of the imaging data and will provide an independent assessment of tumor response data for all patients according to an IRF charter. PFS assessment by IRF will be used for the co-primary endpoint analysis, and other IRF-assessed tumor response data will be examined for analyses of secondary and exploratory efficacy endpoints (Section 4.4.1.2).

4. STATISTICAL METHODS

The analyses described in this SAP will supersede those specified in the YO40245 protocol if there is any inconsistency between the two documents for the purpose of a regulatory filing.

4.1 ANALYSIS POPULATIONS

4.1.1 Intent-to-Treat Population

The ITT population is defined as all randomized patients, whether or not the patient has received the assigned study treatment.

Patients will be grouped according to the treatment assigned at randomization by the interactive voice or web-based response system (IxRS), whether or not the assigned treatment was received.

4.1.2 Pharmacokinetic-Evaluable Population

The pharmacokinetic (PK)-evaluable population is defined as all patients who received any dose of study treatment and who have at least one post-baseline PK sample available.

4.1.3 PRO-Evaluable Population

The patient-reported outcome (PRO)-evaluable population will include all randomized patients who have a baseline and at least 1 post-baseline assessment. The

PRO-evaluable population will be used for descriptive analyses of visit summary and change from baseline and proportion analyses.

The ITT population will be used for the analyses of PRO completion and time-to-deterioration (TTD).

All PRO analyses will be performed per the treatment arm assigned at randomization by the IxRS, whether or not the assigned treatment was received.

4.1.4 Safety Population

The safety population consists of all randomized patients who received at least one full or partial dose of any study treatment, with patients grouped according to the actual treatment received.

Patients who received any amount of atezolizumab and/or bevacizumab will be assigned to treatment Arm A for safety analyses even if atezolizumab/bevacizumab was given in error.

4.1.5 Anti-Drug-Antibody-Evaluable Population

The anti-drug-antibody (ADA)-evaluable population is defined as all patients who received any dose of atezolizumab and who have at least one post-baseline ADA assessment.

4.2 ANALYSIS OF STUDY CONDUCT

Enrollment, major protocol deviations (including major deviations of inclusions/exclusion criteria), and reasons for discontinuation from the study will be summarized by treatment arm for the ITT population. Study treatment administration and reasons for discontinuation from study treatment will be summarized by treatment arm for the safety population.

4.3 ANALYSIS OF TREATMENT GROUP COMPARABILITY

Demographic characteristics such as age, sex, race/ethnicity, baseline disease characteristics (e.g., etiology, disease stage, Child-Pugh classification, prior treatments, etc) and stratification factors will be summarized by treatment arm for the ITT population.

Baseline measurements are the last available data obtained prior to the patient receiving the first dose of any component of study treatment; if a patient did not receive any study treatments, baseline measurements are the last available data obtained prior to randomization date unless otherwise specified.

Medical history will be summarized by arm based on the safety population. The summary will be separately reported for resolved events/conditions versus ongoing events/conditions as collected at baseline. Concomitant medications will also be summarized based on the safety population for medications taken prior to the first dose

of study treatment, regardless of whether medications were ongoing or not post-the start of treatment, and the medications taken after the first dose of study treatment.

Descriptive statistics (mean, median, standard deviation, and range) will be presented for continuous data, and frequencies and percentages will be presented for categorical data.

4.4 EFFICACY ANALYSIS

Unless otherwise specified, efficacy analyses will be conducted based on the ITT population.

Unless otherwise specified, geographic region (Asia excluding Japan vs. rest of world), macrovascular invasion and/or extrahepatic spread (presence vs. absence) and baseline AFP ($<$400 vs. $\geq$400 ng/mL) per IxRS at randomization will be used in all the stratified analysis. These are factors that are considered to be among the most prognostic factors for the patient population indicated in the study. ECOG PS is removed from the stratified analysis to avoid the potential risk of over-stratification (Akazawa et al. 1997).

4.4.1 Co-Primary Efficacy Endpoints

4.4.1.1 Overall Survival (OS)

Overall survival is defined as the time from the date of randomization to the date of death from any cause. Patients who are alive at the time of the analysis data cutoff will be censored at the last date they were known to be alive. Patients with no post-baseline information will be censored at the date of randomization.

The null and alternative hypotheses regarding OS can be phrased in terms of the survival functions $S_A(t)$ and $S_B(t)$ for Arm A (atezolizumab + bevacizumab) and Arm B (sorafenib), respectively:

$$H0: S_{OS\_A}(t) = S_{OS\_B}(t) \text{ versus } H1: S_{OS\_A}(t) \neq S_{OS\_B}(t)$$

The stratified two-sided log-rank test will be used as the primary analysis to compare OS between the two treatment arms. The results from the unstratified log-rank test will also be provided.

The Kaplan-Meier method will be used to estimate median OS for each treatment arm. Brookmeyer-Crowley methodology will be used to calculate the 95% confidence interval (CI) for the median OS for each treatment arm. Stratified Cox proportional-hazards models will be used to estimate the HR and its 95% CIs. The unstratified HR will also be provided.

As described in Section 2.2, if the null hypotheses of the PFS and key secondary endpoints testing are all rejected at a two-sided significance level of 0.002, OS will be tested at an overall two-sided significance level of 0.05. Otherwise, OS will be tested at

an overall two-sided significance level of 0.048. As described in Section 2.3, three analyses for OS are planned, including two interim analyses.

4.4.1.2 PFS by IRF-Assessment (PFS-IRF) per RECIST v1.1

PFS-IRF is defined as the time from randomization to the occurrence of disease progression as determined by IRF according to RECIST v1.1, or death from any cause, whichever occurs first. Patients who have not experienced disease progression or death at the time of the clinical cutoff date will be censored at the time of the last tumor assessment on or prior to the clinical cutoff date. Patients with no post-baseline tumor assessment will be censored at the date of randomization.

The null and alternative hypotheses regarding PFS-IRF can be phrased in terms of the PFS functions $S_A(t)$ and $S_B(t)$ for Arm A (atezolizumab + bevacizumab) and Arm B (sorafenib), respectively:

$$H0: S_{PFS\_A}(t) = S_{PFS\_B}(t) \text{ versus } H1: S_{PFS\_A}(t) \neq S_{PFS\_B}(t)$$

PFS-IRF will be analyzed with the same methodologies as OS. Treatment comparisons will be based on the stratified log-rank test. If the null hypothesis of the OS testing is rejected at a two-sided significance level of 0.048 (as indicated in Table 2), PFS will be tested at the two-sided significance level of 0.05. Otherwise, PFS will be tested at the two-sided significance level of 0.002.

4.4.2 Secondary Efficacy Endpoints

4.4.2.1 Objective Response Rate (ORR)

ORR by IRF-Assessment (ORR-IRF) per RECIST v1.1 and HCC mRECIST

The analysis population for the analysis of objective response rate per IRF (ORR-IRF) according to RECIST v1.1 will be all patients in the ITT population with measurable disease at baseline per IRF according to RECIST v1.1. The analysis population for the analysis of ORR-IRF according to HCC modified RECIST (mRECIST) will be all patients in the ITT population with measurable disease at baseline per IRF according to HCC mRECIST.

An objective response per IRF is defined as a complete response (CR) or partial response (PR) as determined by an IRF, and it is assessed separately according to RECIST v1.1 and HCC mRECIST. Patients not meeting these criteria, including patients without any post-baseline tumor assessment, will be considered non-responders.

ORR-IRF is defined as the proportion of patients who had an objective response as assessed by IRF. Objective responses, confirmation not required and confirmation required (i.e., CR or PR at two consecutive tumor assessments at least 28 days apart), will be separately considered for the ORR analysis.

The two-sided Cochran-Mantel-Haenszel test will be used to compare ORR between the two treatment arms. ORR-IRF is calculated for each treatment arm and the difference in ORR-IRF between treatment arms will be computed. The 95% CI for ORR-IRF for each arm will be derived using the Clopper-Pearson method (Clopper and Pearson 1934). The 95% CI for difference in ORR is computed by normal approximation.

ORR by Investigator-Assessment (ORR-INV) per RECIST v1.1

The analysis population for ORR-investigator (ORR-INV) will be all patients in the ITT population with measurable disease at baseline per INV according to RECIST v1.1.

Investigator-assessed objective response is defined as a CR or PR as determined by the investigator according to RECIST v1.1. Patients not meeting these criteria, including patients without any post-baseline tumor assessment, will be considered non-responders.

ORR-INV is defined as the percentage of patients who have an objective response per the investigator according to RECIST v1.1. Objective responses, confirmation not required, and confirmation required (i.e., CR or PR at two consecutive tumor assessments at least 28 days apart), will be separately considered for the ORR analysis. The analysis methods are similar to those described for ORR-IRF.

4.4.2.2 Duration of Response (DOR)

Duration of response will be assessed in patients who had an objective response (confirmation not required, and confirmation required, analysis will be conducted separately). DOR is defined as the time interval from the date of first occurrence of a documented objective response (CR or PR, whichever status is recorded first) until the first date that disease progression or death is documented, whichever occurs first. Patients who have not progressed and who have not died at the time of analysis will be censored at the time of last tumor assessment. If no tumor assessments were performed after the date of the first occurrence of a documented CR or PR, DOR will be censored at the date of the first occurrence of a documented CR or PR. The analysis of DOR is based on a non-randomized subset of patients (specifically, patients who achieved an objective response); therefore, comparisons between treatment arms will be made for descriptive purposes only.

DOR analyses will be performed separately based on IRF-assessed (DOR-IRF) tumor response according to RECIST v1.1 and HCC mRECIST, and investigator-assessed (DOR-INV) tumor response according to RECIST v1.1. The analysis methods are similar to those described for PFS.

4.4.2.3 Progression-Free Survival (PFS)

PFS analyses will be performed separately based on investigator-assessed tumor response (PFS-INV) according to RECIST v1.1 and IRF-assessed tumor response

(PFS-IRF) according to HCC mRECIST. The analysis methods are similar to those described for OS.

4.4.2.4 Time to Progression (TTP)

Time to progression is defined as the time from the date of randomization to the date of the first documented tumor progression. Patients without tumor progression will be censored at the last tumor assessment date. Patients who have no post-baseline tumor assessment will be censored at the date of randomization.

TTP analyses will be performed separately based on IRF-assessed tumor response (TTP-IRF) according to RECIST v1.1 and HCC mRECIST as well as investigator-assessed tumor response (TTP-INV) according to RECIST v1.1. The analysis methods are similar to those described for OS.

4.4.2.5 PFS and OS by Baseline AFP

PFS-IRF and PFS-INV according to RECIST v1.1 and OS will be analyzed by subgroups of baseline serum AFP level ($<$400 ng/mL vs. $\geq$400 ng/mL). The analysis methods are similar to those described for the co-primary endpoints. Geographic region (Asia excluding Japan vs. rest of the world) and macrovascular invasion and/or extrahepatic spread (presence vs absence) will be used as stratification factors for stratified analyses.

4.4.2.6 Time to Deterioration (TTD)

The primary analysis population for evaluation of TTD will be the ITT population.

TTD is defined as the time from randomization to the first deterioration (decrease from baseline of $\geq$10 points) in the patient-reported health-related global health status/quality of life (GHS /HRQoL), physical function or role function scales of the European Organization for Research and Treatment of Cancer quality-of-life questionnaire for cancer (EORTC) QLQ-C30, maintained for two consecutive assessments, or one assessment followed by death from any cause within 3 weeks. A $\geq$10-point change in the EORTC scale score was perceived by patients as clinically significant (Osoba et al. 1998). The Kaplan-Meier analysis methods for the analysis of TTD are similar to those described for PFS. Patients who do not have an observed deterioration prior to discontinuation from study treatment or initiation of non-protocol anti-cancer therapy (NPT), or at the time of the clinical data cut-off, will be censored at the last available assessment date prior to or at the time of discontinuation from study treatment or initiation of NPT or the clinical cut-off date, whatever is earlier. Patients without a post-baseline assessment will be censored at randomization date.

4.4.2.7 Overall Multiplicity Control for Key Secondary Endpoints

If the co-primary endpoint of PFS-IRF according to RECIST v1.1 is statistically significant, then ORR-IRF (confirmation required) according to RECIST v1.1 and ORR-IRF (confirmation required) according to HCC mRECIST will be hierarchically tested. Specifically, ORR-IRF per RECIST v1.1 will be tested first and if it is statistically

significant, ORR-IRF per HCC mRECIST will then be tested. If ORR-IRF per RECIST v1.1 is not statistically significant, ORR-IRF per HCC mRECIST will not be tested. Implementation of this ordered statistical testing procedure will strongly control the type I error at 5% (two-sided) among all key hypotheses.

4.4.3 Exploratory Efficacy Endpoints

PFS-INV, TTP-INV, ORR-INV, and DOR-INV according to immune-modified RECIST (imRECIST), using the same definitions for these efficacy endpoints as outlined in Sections 4.4.1 and 4.4.2, will be analyzed using similar methodology as specified in Sections 4.4.1 and 4.4.2.

4.4.4 Patient-Reported Outcome (PRO) Analyses

Completion rates and reasons for missing data will be summarized for each of the EORTC QLQ-C30 and EORTC QLQ-HCC18 measures at each cycle by treatment arm.

In addition, the following analyses will be conducted by treatment arm.

Exploratory TTD analyses will be conducted utilizing Kaplan-Meier methods, with TTD defined as the time from randomization to first deterioration (decrease from baseline of $\geq$10 points), maintained for two consecutive timepoints, one timepoint followed by death from any cause within 3 weeks for select subscales of the EORTC QLQ-C30 (appetite loss, diarrhea, fatigue, pain) and EORTC QLQ-HCC18 (jaundice, fatigue, pain). Patients who do not have an observed deterioration prior to discontinuation from study treatment or initiation of NPT, or at the time of the clinical data cut-off, will be censored at the last available assessment date prior to or at the time of discontinuation from study treatment or initiation of NPT or the clinical cut-off date, whatever is earlier. Patients without a post-baseline assessment will be censored at randomization date.

Kaplan-Meier analysis methods similar to those described for OS will be used to assess long-term impact on GHS/HRQoL, physical function and role function. For this exploratory analysis, TTD will be defined as the time from randomization to first deterioration (decrease from baseline of $\geq$ 10 points), maintained for two consecutive timepoints, or one timepoint followed by death from any cause within 3 weeks. Patients who do not have an observed deterioration at the time of the clinical data cut-off will be censored at the last available assessment date. Patients without a post-baseline assessment will be censored at randomization.

Visit mean summary and change from baseline analyses will be performed for all subscales of the EORTC QLQ-C30 and QLQ-HCC18. Summary statistics (number of patients, mean, standard deviation, median, minimum, maximum, 95% CI) of score(s) and score change(s) from baseline to each time point. Previously published minimally important differences (MID) (e.g., 10-point MID) will be used to identify meaningful change from baseline within each treatment group on all scales of the EORTC QLQ-C30 and EORTC QLQ-HCC18 (Osoba et al. 1998).

Proportion of patients with a clinically meaningful change (improved, deteriorated, remained stable) in select scales of the QLQ-C30 (GHS/QoL, physical function, role function, appetite loss, diarrhea, fatigue, pain) and QLQ-HCC18 (jaundice, fatigue, pain) will be summarized by treatment arm.

Proportion of responses for itching item (of HCC18 jaundice subscale) and abdominal pain item (of HCC18 pain subscale) at each timepoint while patients are on treatment will be summarized by treatment arm for each response option.

In the event of incomplete data for any questionnaire subscales, if more than 50% of the constituent items are completed, a prorated score will be computed, consistent with the scoring manuals and validation papers. For subscales with less than 50% of the items completed, the subscale will be considered as missing (Fayers et al. 2001; Chie et al. 2012).

4.4.5 Exploratory Biomarker Analysis

Exploratory biomarker analyses may be performed in an effort to understand the association of tissue or blood-based biomarkers with response to atezolizumab + bevacizumab versus sorafenib, or increase the understanding of HCC disease evolution under atezolizumab + bevacizumab treatment and may not be included in the CSR. PFS and OS will be analyzed using the methods outlined in Section 4.4.1.

PFS-IRF according to RECIST v 1.1 and OS by the following biomarkers in tumor tissue may be analyzed:

- Baseline programmed death-ligand 1 (PD-L1) protein expression in tumor tissue
- Baseline expression of T effector gene signature in tumor tissue
- CD8 protein expression level or CD8+ T cell localization
- IHCs or genes/gene signatures (by gene expression profiling) related to tumor microenvironments.

4.4.6 Sensitivity Analyses

Impact of Missing Scheduled Tumor Assessments on Co-Primary Endpoint PFS-IRF per RECIST v1.1

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

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

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

Additional sensitivity analyses may be conducted.

4.4.7 Subgroup Analyses

In order to assess the consistency of treatment effect with respect to the co-primary efficacy endpoints of PFS-IRF according to RECIST v 1.1 and OS across important subgroups, forest plots (including estimated HRs) will be provided, including, but not limited to, the following variables: age, sex, race, geographic region, macrovascular invasion and/or extrahepatic spread, macrovascular invasion, extrahepatic spread, ECOG performance status, HCC etiology, BCLC staging at the time of study entry and baseline PD-L1 expression in tumor tissue for patients with baseline tumor samples. Unstratified analysis results will be presented for subgroup analyses due to the potentially limited number of patients in each subgroup.

To assess how patient baseline characteristics such as invasion of the bile duct, invasion at the main portal vein, and liver occupation impacts the clinical activity of the combination treatment of atezolizumab + bevacizumab, further exploratory efficacy subgroup analyses for the co-primary endpoints of PFS-IRF per RECIST v1.1 and OS may be performed for ITT patients who do not have obvious invasion of the bile duct, invasion at the main portal vein, or < 50% liver occupation.

4.5 PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES

Pharmacokinetic analyses will be performed for the pharmacokinetic-evaluable population.

Samples will be collected for PK analyses and to compare exposure in this study with that attained in previous studies. Serum concentrations of atezolizumab will be reported as individual values and summarized (mean, standard deviation, coefficient of variation, median, range, geometric mean, and geometric mean coefficient of variation) by treatment arm and cycle, when appropriate and as data allow. Individual and median serum atezolizumab concentrations will be plotted by treatment arm and day.

Atezolizumab concentration data may be pooled with data from other studies using an established population PK model to derive PK parameters such as clearance, volume of distribution, and area under the curve, as warranted by the data. Potential correlations of relevant PK parameters with dose, safety, efficacy, or biomarker outcomes may be explored.

Additional PK and pharmacodynamic analyses will be conducted as appropriate. These additional PK analyses may not be included in the CSR.

4.6 SAFETY ANALYSES

Unless specified otherwise, safety analyses described in this section will be conducted based on the safety population.

4.6.1 Exposure of Study Medication

Study drug exposure status, which includes treatment duration, number of cycles, and dose intensity, will be summarized for each treatment arm with descriptive statistics.

4.6.2 Adverse Events

Verbatim description of adverse events will be mapped to Medical Dictionary for Regulatory Activities (MedDRA) thesaurus terms and graded by the investigator according to National Cancer Institute Common Terminology Criteria for Adverse Events, Version 4.0 (NCI CTCAE v4.0). Treatment-emergent adverse events will be summarized by mapped term, appropriate thesaurus level, NCI CTCAE grade, and treatment arm. For reporting purposes, "treatment-emergent" is defined as all AEs with an onset on or after the first exposure to the study treatment, or have AE intensity worsening during the treatment if onset was prior to the first exposure, until the clinical cutoff date. In addition, serious adverse events, severe adverse events (Grade ≥ 3), atezolizumab adverse events of special interest (AESIs), bevacizumab AESIs, immune-mediated AEs (imAEs [defined as atezolizumab AESIs requiring the use of systemic corticosteroids]) and adverse events leading to study drug discontinuation and AEs leading to study drug interruption/modification will be summarized accordingly. Multiple occurrences of the same event will be counted once at the maximum grade.

All deaths and causes of deaths reported during the study treatment period and the follow-up period after treatment completion and/or discontinuation will be summarized by treatment arm.

4.6.3 Laboratory Data

Laboratory data will be summarized by treatment arm. Values outside the normal ranges will be summarized by treatment arm. In addition, selected laboratory data will be summarized by treatment arm and NCI CTCAE grade according to NCI CTCAE v4.0.

To evaluate if there is any potential risks of drug-induced liver injury (DILI) in HCC patients, a listing for all cases meeting the potential Hy's law criteria, defined as total bilirubin increase by $> 2 \times$ ULN within 7 days after ALT or ASL increase by $3 \times$ baseline value, will be provided.

4.6.4 Vital Signs

Changes in selected vital signs will be summarized by treatment arm and by change over time, which includes change from baseline. Baseline is defined as the measurement obtained on Cycle 1, Day 1 before the first dose of study drug is administered.

4.6.5 Electrocardiograms

Post-baseline abnormal electrocardiogram (ECG) (12-lead) records over time will be listed by treatment arm.

4.7 IMMUNOGENICITY ANALYSES

The immunogenicity analysis population will consist of all patients with at least one post-baseline ADA assessment. Patients will be grouped according to treatment received or, if no treatment is received prior to study discontinuation, according to treatment assigned.

The numbers and proportions of ADA-positive patients and ADA-negative patients at baseline (baseline prevalence) and after drug administration (post-baseline incidence) will be summarized by treatment group.

In the ADA-evaluable population, patients are considered as treatment-emergent ADA positive if they were ADA negative at baseline and ADA positive at any time point post baseline (treatment induced), or if they were ADA positive at baseline and had a post baseline ADA positive titer that was 0.6 titer units higher than baseline (treatment enhanced). Otherwise, all other ADA-evaluable patients are considered as treatment emergent ADA negative.

The safety, efficacy, PK, and biomarker endpoints by treatment-emergent ADA status may be analyzed and reported.

4.8 EXPLORATORY HEALTH STATUS ANALYSIS

Health utility data from the EuroQol 5 Dimension Questionnaire, 5-level version (EQ-5D-5L) will be evaluated in pharmacoeconomic models. The results from the health economic data analyses will be reported separately from the CSR.

4.9 MISSING DATA

See Section 4.4.4 and Section 4.4.5 for methods for handling missing data for the primary and secondary endpoints.

4.10 INTERIM ANALYSES

Please refer to Section 2.3 for details on the planned interim analyses.

5. CHINA SUBPOPULATION ANALYSIS

The Sponsor is targeting a total enrollment of approximately 135 patients from mainland China. The sample size of the China subpopulation was considered adequate to characterize the efficacy and safety profile of atezolizumab + bevacizumab in Chinese patients. After approximately 480 patients have been randomized into the global portion of the study, in the event that fewer than 135 patients from mainland China are enrolled, additional patients in China may be subsequently randomized into the two treatment

arms in a 2:1 ratio in an extended China enrollment phase to ensure a total of approximately 135 mainland China patients for the China subpopulation. The same randomization method will be implemented for the China extension cohort in the event the extension cohort opens.

The primary efficacy objective of the China subpopulation analysis is to assess efficacy, as measured by the co-primary endpoints of PFS by an IRF assessment per RECIST v1.1 and OS, of atezolizumab + bevacizumab compared with sorafenib in Chinese patients. The China subpopulation is not powered to demonstrate statistical significance in terms of efficacy, and no formal hypothesis testing will be performed.

The PFS analysis for the China subpopulation will be conducted at the time of the primary PFS analysis for the global population (anticipated to occur approximately 16 months from global FPI), at which point it is estimated that there will be a sufficient number of PFS events to demonstrate at least 80% probability of maintaining 50% of PFS risk reduction compared with that estimated from the global population. If the PFS data in the China subpopulation is not mature at the time of primary analysis, an additional PFS analysis in the China subpopulation may be conducted.

The OS data from the China subpopulation will not be mature at the time of this PFS analysis. Therefore, the Sponsor may decide to conduct updated OS analyses for the China subpopulation at the time of the 2nd OS IA for global population, or until there is a sufficient number of OS events to achieve at least 80% probability of maintaining 50% OS risk reduction compared with that estimated for the global population.

The analysis methods for the China subpopulation will be the same as for the global population unless otherwise noted. Given that all patients in China subpopulation are from geographic region Asia excluding Japan, only macrovascular invasion and/or extrahepatic spread (presence vs. absence) and baseline AFP ($<$400 vs. $\geq$400 ng/mL) will be used as stratification factors in stratified analyses conducted for the China subpopulation. The results from the unstratified analyses will also be provided. The results of the China subpopulation analyses will be summarized in a separate report from the CSR for the global population.

REFERENCES

Akazawa K, Nakamura T, Palesch Y. Power of Logrank Test and Cox Regression Model in Clinical Trials with Heterogeneous Samples. Statistics in Medicine 1997;16:583-97.

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

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

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

Chie WC, Blazeby JM, Hsiao CF, et al. International cross-cultural field validation of an European Organization for Research and Treatment of Cancer questionnaire module for patients with primary liver cancer, the European Organization for Research and Treatment of Cancer quality-of-life questionnaire HCC18. Hepatology 2012;55:1122–9.

Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 1934;26:404−13.

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

Fayers PM, Aaronson NK, Bjordal K, et al. The EORTC QLQ-C30 scoring manual, 3rd ed. Brussels: European Organisation for Research and Treatment of Cancer, 2001.

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

Appendix 1

Protocol Synopsis

PROTOCOL SYNOPSIS

TITLE:	**A PHASE III, OPEN-LABEL, RANDOMIZED STUDY OF ATEZOLIZUMAB IN COMBINATION WITH BEVACIZUMAB COMPARED WITH SORAFENIB IN PATIENTS WITH UNTREATED LOCALLY ADVANCED OR METASTATIC HEPATOCELLULAR CARCINOMA**
PROTOCOL NUMBER:	YO40245
VERSION NUMBER:	4
EUDRACT NUMBER:	2017-003691-31
IND NUMBER:	135913
TEST PRODUCTS:	Atezolizumab (RO5541267) Bevacizumab (RO4876646)
PHASE:	III
INDICATION:	Metastatic hepatocellular carcinoma (HCC)
SPONSOR:	F. Hoffmann-La Roche Ltd

Objectives and Endpoints

This study will evaluate the efficacy and safety of atezolizumab in combination with bevacizumab compared with sorafenib in patients with locally advanced or metastatic HCC who have received no prior systemic treatment. Specific objectives and corresponding endpoints for the study are outlined below.

Primary Efficacy Objective	Corresponding Endpoint
• To evaluate the efficacy of atezolizumab + bevacizumab compared with sorafenib	• OS, defined as the time from randomization to death from any cause • PFS, defined as the time from randomization to the first occurrence of disease progression or death from any cause (whichever occurs first), as determined by an IRF according to RECIST v1.1

Secondary Efficacy Objectives	Corresponding Endpoints
• To evaluate the efficacy of atezolizumab + bevacizumab compared with sorafenib	• Objective response, defined as a complete or partial response, as determined by the investigator according to RECIST v1.1 • PFS as determined by the investigator according to RECIST v1.1 • TTP, defined as the time from randomization to the first occurrence of disease progression, as determined by the investigator according to RECIST v1.1 • DOR, defined as the time from the first occurrence of a documented objective response to disease progression or death from any cause (whichever occurs first), as determined by the investigator according to RECIST v1.1 • Objective response as determined by an IRF according to RECIST v1.1 • TTP as determined by an IRF according to RECIST v1.1 • DOR as determined by an IRF according to RECIST v1.1 • Objective response as determined by an IRF according to HCC mRECIST • PFS as determined by an IRF according to HCC mRECIST • TTP as determined by an IRF according to HCC mRECIST • DOR as determined by an IRF according to HCC mRECIST
• To evaluate the association of pre-specified biomarkers with efficacy of atezolizumab + bevacizumab compared with sorafenib	• PFS as determined by the investigator and by an IRF according to RECIST v1.1 and OS by baseline serum AFP level ($<$ 400 ng/mL vs. $\geq$ 400 ng/mL)
• To evaluate PROs of disease/treatment-related symptoms, GHS/QoL, and function experienced by patients on atezolizumab + bevacizumab versus sorafenib	• TTD, defined as the time from randomization to first deterioration (decrease from baseline of $\geq$ 10 points), maintained for two consecutive $assessments$ or one $assessment$ followed by death from any cause within 3 weeks in the following EORTC QLQ-C30 subscales: – Physical functioning (PF) – Role functioning (RF) – GHS/QoL
Exploratory Efficacy Objectives	**Corresponding Endpoints**
• To evaluate the efficacy of atezolizumab + bevacizumab compared with sorafenib	• Objective response as determined by the investigator according to imRECIST • PFS as determined by the investigator according to imRECIST • TTP as determined by the investigator according to imRECIST • DOR as determined by the investigator according to imRECIST

Exploratory Efficacy Objectives (cont.)	Corresponding Endpoints
• To evaluate PROs of disease/treatment-related symptoms (including abdominal pain and itching), GHS/QoL, and function experienced by patients on atezolizumab + bevacizumab versus sorafenib	• Mean and mean changes from baseline score (by cycle) in all the subscales of the EORTC QLQ-C30 and QLQ-HCC18 • Proportion of patients with clinically meaningful change in select scales of the QLQ-C30 *(GHS/QoL, physical function, role function, appetite loss, diarrhea, fatigue, pain) and QLQ-HCC18 (fatigue, jaundice, pain)* • TTD maintained for two consecutive timepoints, or one timepoint followed by death from any cause within 3 weeks*, in select scales of the QLQ-C30 (appetite loss, diarrhea, fatigue, pain) and QLQ-HCC18 (jaundice, fatigue, pain)* • *TTD maintained for two consecutive timepoints, or one timepoint followed by death from any cause, within 3 weeks in HRQoL/GHS, physical function, and role function of the QLQ-C30* • Proportion of responses for abdominal pain item (of QLQ-HCC18 pain subscale) and itching item (of QLQ-HCC18 jaundice subscale)
Safety Objective	**Corresponding Endpoints**
• To evaluate the safety of atezolizumab + bevacizumab compared with sorafenib	• Incidence and severity of adverse events, with severity determined according to NCI CTCAE v4.0 • Vital signs • Clinical laboratory test results
Pharmacokinetic Objective	**Corresponding Endpoint**
• To characterize the PK profile of atezolizumab when given in combination with bevacizumab	• Serum concentration of atezolizumab at specified timepoints
Immunogenicity Objective	**Corresponding Endpoint**
• To evaluate the immune response to atezolizumab	• Presence of ADAs to atezolizumab during the study relative to the presence of ADAs at baseline
Exploratory Immunogenicity Objective	**Corresponding Endpoint**
• To evaluate potential effects of ADAs to atezolizumab on the efficacy, safety, and pharmacokinetics of atezolizumab + bevacizumab	• Efficacy, safety, or PK endpoints by ADA status

Exploratory Biomarker Objective	Corresponding Endpoint
• To identify tissue or blood based biomarkers that are associated with response to atezolizumab + bevacizumab versus sorafenib, or can increase the understanding of HCC disease evolution under atezolizumab + bevacizumab treatment	• PFS as determined by the investigator according to RECIST v 1.1 and OS based on the following biomarkers in tumor tissue: – Baseline expression of T effector gene signature in tumor tissue – Baseline PD-L1 protein expression in tumor tissue – CD8 protein expression level or CD8+ T cell localization – IHCs or genes/gene signatures (by gene expression profiling) related to tumor microenvironments – T cell receptor sequence profile in tumor-associated T cells • Objective response as determined by the investigator according to RECIST v1.1 and OS based on the following biomarkers in blood: – Immune-related biomarkers profiling in plasma and serum (e.g., interleukin 2, interferon-γ)
Exploratory Health Status Objective	**Corresponding Endpoint**
• To evaluate health status experienced by patients on atezolizumab + bevacizumab versus sorafenib to generate utility scores for use in economic evaluations	• Health utility and VAS scores of the EQ-5D-5L questionnaire

ADA = anti-drug antibody; AFP = α-fetoprotein; DOR = duration of response; EORTC = European Organisation for Research and Treatment of Cancer; GHS = global health status; EQ-5D-5L = EuroQol 5-Dimension Questionnaire, 5-level version; HCC = hepatocellular carcinoma; HCC mRECIST = modified RECIST for HCC; HRQoL = health-related quality of life; imRECIST = immune-modified RECIST; IRF = independent review facility; NCI CTCAE v4.0 = National Cancer Institute Common Terminology Criteria for Adverse Events, Version 4.0; OS = overall survival; PFS = progression-free survival; PRO = patient-reported outcome; PK = pharmacokinetic; QLQ-C30 = quality-of-life questionnaire for cancer; QLQ-HCC18 = HCC disease-specific module; QoL=quality of life; RECIST v1.1 = Response Evaluation Criteria in Solid Tumors, Version 1.1; TTD = time to deterioration; TTP = time to progression; VAS = visual analog scale.

Study Design

Description of Study

This is a Phase III, randomized, multicenter, open-label, two-arm study designed to evaluate the efficacy and safety of atezolizumab + bevacizumab versus sorafenib in patients with locally advanced or metastatic HCC who have received no prior systemic treatment.

This study will enroll approximately 480 patients randomized in a 2:1 ratio to one of two treatment arms:

- Arm A (experimental arm): Atezolizumab 1200 mg IV infusions Q3W (dosed in 3-week cycles) + bevacizumab 15 mg/kg Q3W (dosed in 3-week cycles)
- Arm B (control arm): Sorafenib 400 mg by mouth (PO), twice per day (BID), continuously

Randomization will be stratified according to the following stratification factors:

- Geographic region (Asia excluding Japan vs. rest of world)
- Macrovascular invasion and/or extrahepatic spread (presence vs. absence)
- Baseline AFP (< 400 vs. ≥ 400 ng/mL)
- ECOG performance status (0 vs. 1)

Patients randomized to the atezolizumab + bevacizumab arm (Arm A) who transiently withhold or permanently discontinue either atezolizumab or bevacizumab may continue on single-agent therapy as long as the patients are experiencing clinical benefit in the opinion of the investigator

and after discussion with the Medical Monitor (i.e., patients transiently withhold or permanently discontinue bevacizumab for adverse effects may continue atezolizumab monotherapy and vice versa).

Patients will receive atezolizumab and/or bevacizumab or sorafenib until unacceptable toxicity or loss of clinical benefit as determined by the investigator after an integrated assessment of radiographic and biochemical data, and clinical status (e.g., symptomatic deterioration such as pain secondary to disease). In the absence of unacceptable toxicity, patients who meet criteria for disease progression per RECIST v1.1 while receiving atezolizumab and/or bevacizumab or sorafenib will be permitted to continue the study treatment if they meet all of the following criteria:

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

Tumor assessments will be performed at baseline and at regular intervals during study treatment. Additional scans will be performed as clinically indicated. Tumor assessments will continue until disease progression, regardless of whether treatment has been discontinued (e.g., for toxicity). Patients who meet RECIST v1.1 criteria for progression will undergo tumor assessments until disease progression (per immune-modified RECIST [imRECIST]) or loss of clinical benefit, whichever occurs later. In the absence of disease progression, tumor assessments should continue regardless of whether patients start new anti-cancer therapy, until consent is withdrawn, death, or the study is terminated by the Sponsor, whichever occurs first. Following disease progression, patients will be followed for survival and subsequent anti-cancer therapies until death, loss to follow-up, withdrawal of consent, or study termination by Sponsor, whichever occurs first.

Sites will provide imaging used for tumor assessments to an IRF to enable centralized, independent review of response and progression endpoints. These reviews will be performed prior to the pre-specified efficacy analyses. IRF membership and procedures will be detailed in an IRF Charter.

Safety assessments will include the incidence, nature, and severity of adverse events and laboratory abnormalities graded per the National Cancer Institute Common Terminology Criteria for Adverse Events, Version 4.0 (NCI CTCAE v4.0). Laboratory safety assessments will include the regular monitoring of hematology and blood chemistry. Serum samples will be collected to monitor the pharmacokinetics of atezolizumab when administered in combination with bevacizumab. Patient samples, including archival tumor tissues as well as serum and plasma, will be collected for future exploratory biomarker assessments.

Number of Patients

This study will initially enroll approximately 480 patients across all sites in a global enrollment phase. The final analysis of the global study will be based on the global population, which will include all patients enrolled during the global enrollment phase (including patients enrolled in China during that phase).

The Sponsor is targeting a total enrollment of approximately 135 patients from mainland China. After completion of the global enrollment phase, in the event that less than 135 patients from mainland China are enrolled, additional patients in China may be subsequently randomized into the two treatment arms in a 2:1 ratio in an extended China enrollment phase to ensure a total of approximately 135 patients *from mainland China* in a China subpopulation. *The* China subpopulation will include all patients enrolled in China *(i.e., during both the global enrollment phase and the extended China enrollment phase). The patients enrolled in the China extension phase will undergo the same schedule of activities and will receive atezolizumab + bevacizumab, or sorafenib as in the global study.* The China subgroup analysis will be performed based on the China subpopulation.

Target Population

Inclusion Criteria

Patients must meet the following criteria for study entry:

- Signed Informed Consent Form
- Age $\geq$ 18 years at time of signing Informed Consent Form
- Ability to comply with the study protocol, in the investigator's judgment
- Locally advanced or metastatic and/or unresectable HCC with diagnosis confirmed by histology/ cytology or clinically by AASLD criteria in cirrhotic patients

  Patients without cirrhosis require histological confirmation of diagnosis.
- Disease that is not amenable to curative surgical and/or locoregional therapies, or progressive disease after surgical and /or locoregional therapies
- No prior systemic therapy (including systemic investigational agents) for HCC

  Previous use of herbal therapies/traditional Chinese medicines with anti-cancer activity included in the label is allowed, provided that these medications are discontinued prior to randomization.
- At least one measurable (per RECIST 1.1) untreated lesion
- Patients who received prior local therapy (e.g., radiofrequency ablation, percutaneous ethanol or acetic acid injection, cryoablation, high-intensity focused ultrasound, transarterial chemoembolization, transarterial embolization, etc.) are eligible provided the target lesion(s) have not been previously treated with local therapy or the target lesion(s) within the field of local therapy have subsequently progressed in accordance with RECIST version 1.1.
- Pre-treatment tumor tissue sample (if available)

  If tumor tissue is available, a formalin-fixed, paraffin-embedded (FFPE) tumor specimen in a paraffin block (preferred) or approximately 10–15 slides containing unstained, freshly cut, serial sections should be submitted along with an associated pathology report within 4 weeks of randomization.

  If FFPE specimens described above are not available, any type of specimens (including fine-needle aspiration, cell pellet specimens [e.g., from pleural effusion], and lavage samples) are also acceptable. This specimen should be accompanied by the associated pathology report.

  If tumor tissue is not available (e.g., depleted because of prior diagnostic testing), patients are still eligible.
- ECOG Performance Status of 0 or 1 within 7 days prior to randomization
- Child-Pugh class A within 7 days prior to randomization
- Adequate hematologic and end-organ function, defined by the following laboratory test results, obtained within 7 days prior to randomization, unless otherwise specified:
  - ANC $\geq 1.5 \times 10^9$/L (1500/μL) without granulocyte colony-stimulating factor support
  - Lymphocyte count $\geq 0.5 \times 10^9$/L (500/μL)
  - Platelet count $\geq 75 \times 10^9$/L (75,000/μL) without transfusion
  - Hemoglobin $\geq$ 90 g/L (9 g/dL)

    Patients may be transfused to meet this criterion.
  - AST, ALT, and alkaline phosphatase (ALP) $\leq 5 \times$ upper limit of normal (ULN)
  - Serum bilirubin $\leq 3 \times$ ULN
  - Serum creatinine $\leq 1.5 \times$ ULN or creatinine clearance $\geq$ 50 mL/min (calculated using the Cockcroft-Gault formula)
  - Serum albumin $\geq$ 28 g/L (2.8 g/dL) without transfusion
  - For patients not receiving therapeutic anticoagulation: INR or aPTT $\leq 2 \times$ ULN

- Urine dipstick for proteinuria $< 2+$ (within 7 days prior to initiation of study treatment)

  Patients discovered to have $\geq 2+$ proteinuria on dipstick urinalysis at baseline should undergo a 24-hour urine collection and must demonstrate < 1 g of protein in 24 hours.

- Resolution of any acute, clinically significant treatment-related toxicity from prior therapy to Grade ≤ 1 prior to study entry, with the exception of alopecia
- Negative HIV test at screening
- Documented virology status of hepatitis, as confirmed by screening HBV and HCV serology test
- For patients with active hepatitis B virus (HBV):

  HBV DNA < 500 IU/mL obtained within 28 days prior to initiation of study treatment, and

  Anti-HBV treatment (per local standard of care; e.g., entecavir) for a minimum of 14 days prior to study entry and willingness to continue treatment for the length of the study
- For women of childbearing potential: agreement to remain abstinent (refrain from heterosexual intercourse) or use contraceptive methods with a failure rate of $< 1\%$ per year during the treatment period and for at least 5 months after the last dose of atezolizumab, 6 months after the last dose of bevacizumab, or 1 month after the last dose of sorafenib. Women must refrain from donating eggs during this same period.

  A woman is considered to be of childbearing potential if she is postmenarcheal, has not reached a postmenopausal state (≥ 12 continuous months of amenorrhea with no identified cause other than menopause), and has not undergone surgical sterilization (removal of ovaries and/or uterus).

  Examples of contraceptive methods with a failure rate of $< 1\%$ per year include bilateral tubal ligation, male sterilization, hormonal contraceptives that inhibit ovulation, hormone-releasing intrauterine devices, and copper intrauterine devices.

  The reliability of sexual abstinence should be evaluated in relation to the duration of the clinical trial and the preferred and usual lifestyle of the patient. Periodic abstinence (e.g., calendar, ovulation, symptothermal, or postovulation methods) and withdrawal are not acceptable methods of contraception.
- For men: agreement to remain abstinent (refrain from heterosexual intercourse) or use contraceptive measures, and agreement to refrain from donating sperm, as defined below:

  With female partners of childbearing potential, men must remain abstinent or use a condom plus an additional contraceptive method that together result in a failure rate of $< 1\%$ per year during the treatment period and for 6 months after the last dose of bevacizumab or 3 months after the last dose of sorafenib. Men must refrain from donating sperm during this same period.

  With pregnant female partners, men must remain abstinent or use a condom during the treatment period and for 6 months after the last dose of bevacizumab or 3 months after the last dose of sorafenib to avoid exposing the embryo.

  The reliability of sexual abstinence should be evaluated in relation to the duration of the clinical trial and the preferred and usual lifestyle of the patient. Periodic abstinence (e.g., calendar, ovulation, symptothermal, or postovulation methods) and withdrawal are not acceptable methods of contraception.
- For the extended China enrollment phase: Chinese ancestry and residence in Mainland China, Hong Kong, or Taiwan with enrollment at sites recognized by the China FDA

Exclusion Criteria

Patients who meet any of the following criteria will be excluded from study entry:

- History of leptomeningeal disease

- Active or history of autoimmune disease or immune deficiency, including, but not limited to, myasthenia gravis, myositis, autoimmune hepatitis, systemic lupus erythematosus, rheumatoid arthritis, inflammatory bowel disease, antiphospholipid antibody syndrome, Wegener granulomatosis, Sjögren syndrome, Guillain-Barré syndrome, or multiple sclerosis, with the following exceptions:

  Patients with a history of autoimmune-related hypothyroidism who are on thyroid-replacement hormone are eligible for the study.

  Patients with controlled Type 1 diabetes mellitus who are on an insulin regimen are eligible for the study.

  Patients with eczema, psoriasis, lichen simplex chronicus, or vitiligo with dermatologic manifestations only (e.g., patients with psoriatic arthritis are excluded) are eligible for the study provided all of following conditions are met:

  – Rash must cover $<10\%$ of body surface area
  – Disease is well controlled at baseline and requires only low-potency topical corticosteroids
  – No occurrence of acute exacerbations of the underlying condition requiring psoralen plus ultraviolet A radiation, methotrexate, retinoids, biologic agents, oral calcineurin inhibitors, or high-potency or oral corticosteroids within the previous 12 months
- History of idiopathic pulmonary fibrosis, organizing pneumonia (e.g., bronchiolitis obliterans), drug-induced pneumonitis, or idiopathic pneumonitis, or evidence of active pneumonitis on screening chest computed tomography (CT) scan

  History of radiation pneumonitis in the radiation field (fibrosis) is permitted.
- Known active tuberculosis
- Significant cardiovascular disease (such as New York Heart Association Class II or greater cardiac disease, myocardial infarction, or cerebrovascular accident within 3 months prior to initiation of study treatment), unstable arrhythmia, or unstable angina
- History of congenital long QT syndrome or corrected QT interval >500 ms (calculated with use of the Fridericia method) at screening
- History of uncorrectable electrolyte disorder affecting serum levels of potassium, calcium, or magnesium
- Major surgical procedure, other than for diagnosis, within 4 weeks prior to initiation of study treatment, or anticipation of need for a major surgical procedure during the study
- History of malignancy other than HCC within 5 years prior to screening, with the exception of malignancies with a negligible risk of metastasis or death (e.g., 5-year OS rate $>90\%$), such as adequately treated carcinoma in situ of the cervix, non-melanoma skin carcinoma, localized prostate cancer, ductal carcinoma in situ, or Stage I uterine cancer
- Severe infection within 4 weeks prior to initiation of study treatment, including, but not limited to, hospitalization for complications of infection, bacteremia, or severe pneumonia
- Treatment with therapeutic oral or IV antibiotics within 2 weeks prior to initiation of study treatment

  Patients receiving prophylactic antibiotics (e.g., to prevent a urinary tract infection or chronic obstructive pulmonary disease exacerbation) are eligible for the study.
- Prior allogeneic stem cell or solid organ transplantation
- Any other disease, metabolic dysfunction, physical examination finding, or clinical laboratory finding that contraindicates the use of an investigational drug, may affect the interpretation of the results, or may render the patient at high risk from treatment complications
- Treatment with a live, attenuated vaccine within 4 weeks prior to initiation of study treatment, or anticipation of need for such a vaccine during atezolizumab treatment or within 5 months after the last dose of atezolizumab
- History of severe allergic anaphylactic reactions to chimeric or humanized antibodies or fusion proteins

- Known hypersensitivity to Chinese hamster ovary cell products or to any component of the atezolizumab or bevacizumab formulation
- Pregnancy or breastfeeding, or intention of becoming pregnant during study treatment or within at least 5 months after the last dose of atezolizumab, 6 months after the last dose of bevacizumab, or 1 month after the last dose of sorafenib

  Women of childbearing potential must have a negative serum pregnancy test result within 14 days prior to initiation of study treatment.
- Known fibrolamellar HCC, sarcomatoid HCC, or mixed cholangiocarcinoma and HCC
- Untreated or incompletely treated esophageal and/or gastric varices with bleeding or high risk for bleeding

  Patients must undergo an esophagogastroduodenoscopy (EGD), and all size of varices (small to large) must be assessed and treated per local standard of care prior to enrollment. Patients who have undergone an EGD within 6 months of prior to initiation of study treatment do not need to repeat the procedure.
- A prior bleeding event due to esophageal and/or gastric varices within 6 months prior to initiation of study treatment
- Moderate or severe ascites
- History of hepatic encephalopathy
- Co-infection of HBV and HCV

  Patients with a history of HCV infection but who are negative for HCV RNA by PCR will be considered non-infected with HCV.
- Symptomatic, untreated, or actively progressing central nervous system (CNS) metastases

  Asymptomatic patients with treated CNS lesions are eligible, provided that all of the following criteria are met:
  - Measurable disease, per RECIST v1.1, must be present outside the CNS.
  - The patient has no history of intracranial hemorrhage or spinal cord hemorrhage.
  - Metastases are limited to the cerebellum or the supratentorial region (i.e., no metastases to the midbrain, pons, medulla, or spinal cord).
  - There is no evidence of interim progression between completion of CNS-directed therapy and initiation of study treatment.
  - The patient has not undergone stereotactic, whole-brain radiotherapy, and/or neurosurgical resection within 28 days prior to initiation of study treatment.
  - The patient has no ongoing requirement for corticosteroids as therapy for CNS disease. Anticonvulsant therapy at a stable dose is permitted.

  Asymptomatic patients with CNS metastases newly detected at screening are eligible for the study after receiving radiotherapy or surgery, with no need to repeat the screening brain scan.
- Uncontrolled tumor-related pain

  Patients requiring pain medication must be on a stable regimen at study entry.

  Symptomatic lesions (e.g., bone metastases or metastases causing nerve impingement) amenable to palliative radiotherapy should be treated prior to enrollment. Patients should be recovered from the effects of radiation. There is no required minimum recovery period.

  Asymptomatic metastatic lesions that would likely cause functional deficits or intractable pain with further growth (e.g., epidural metastasis that is not currently associated with spinal cord compression) should be considered for loco-regional therapy if appropriate prior to enrollment.
- Uncontrolled pleural effusion, pericardial effusion, or ascites requiring recurrent drainage procedures (once monthly or more frequently)

  Patients with indwelling catheters (e.g., PleurX®) are allowed.

- Uncontrolled or symptomatic hypercalcemia (ionized calcium $>$ 1.5 mmol/L, calcium $>$ 12 mg/dL or corrected serum calcium $>$ ULN)
- Treatment with investigational therapy within 28 days prior to initiation of study treatment
- Treatment with strong CYP3A4 inducers within 14 days prior to initiation of study treatment, including rifampin (and its analogues) or St. John's wort
- Prior treatment with CD137 agonists or immune checkpoint blockade therapies, including anti–CTLA-4, anti–PD-1, and anti–PD-L1 therapeutic antibodies
- Treatment with systemic immunostimulatory agents (including, but not limited to, interferon and interleukin 2 [IL-2]) within 4 weeks or 5 half-lives of the drug (whichever is longer) prior to initiation of study treatment
- Treatment with systemic immunosuppressive medication (including, but not limited to, corticosteroids, cyclophosphamide, azathioprine, methotrexate, thalidomide, and anti–TNF-α agents) within 2 weeks prior to initiation of study treatment, or anticipation of need for systemic immunosuppressive medication during study treatment, with the following exceptions:

  Patients who received acute, low-dose systemic immunosuppressant medication or a one-time pulse dose of systemic immunosuppressant medication (e.g., 48 hours of corticosteroids for a contrast allergy) are eligible for the study after Medical Monitor approval has been obtained.

  Patients who received mineralocorticoids (e.g., fludrocortisone), corticosteroids for chronic obstructive pulmonary disease (COPD) or asthma, or low-dose corticosteroids for orthostatic hypotension or adrenal insufficiency are eligible for the study.
- Inadequately controlled arterial hypertension (defined as systolic blood pressure (BP) $\geq$ 150 mmHg and/or diastolic blood pressure $>$ 100 mmHg), based on an average of $\geq$ 3 BP readings on $\geq$ 2 sessions

  Anti-hypertensive therapy to achieve these parameters is allowable.
- Prior history of hypertensive crisis or hypertensive encephalopathy
- Significant vascular disease (e.g., aortic aneurysm requiring surgical repair or recent peripheral arterial thrombosis) within 6 months prior to initiation of study treatment
- History of hemoptysis ($\geq$ 2.5 mL of bright red blood per episode) within 1 month prior to initiation of study treatment
- Evidence of bleeding diathesis or significant coagulopathy (in the absence of therapeutic anticoagulation)
- Current or recent (within 10 days of first dose of study treatment) use of aspirin ($>$ 325 mg/day) or treatment with dipyramidole, ticlopidine, clopidogrel, and cilostazol
- Current or recent (within 10 days prior to study treatment start) use of full-dose oral or parenteral anticoagulants or thrombolytic agents for therapeutic (as opposed to prophylactic) purpose

  Prophylactic anticoagulation for the patency of venous access devices is allowed provided the activity of the agent results in an INR $<$ 1.5 $\times$ ULN and aPTT is within normal limits within 14 days prior to initiation of study treatment.

  For prophylactic use of anticoagulants or thrombolytic therapies, local label approved dose levels may be used.
- Core biopsy or other minor surgical procedure, excluding placement of a vascular access device, within 3 days prior to the first dose of bevacizumab
- History of abdominal or tracheoesophageal fistula, gastrointestinal (GI) perforation, or intra-abdominal abscess within 6 months prior to initiation of study treatment
- History of intestinal obstruction and/or clinical signs or symptoms of GI obstruction including sub-occlusive disease related to the underlying disease or requirement for routine parenteral hydration, parenteral nutrition, or tube feeding prior to initiation of study treatment

Patients with signs/symptoms of sub-/occlusive syndrome/intestinal obstruction at time of initial diagnosis may be enrolled if they had received definitive (surgical) treatment for symptom resolution.

- Evidence of abdominal free air that is not explained by paracentesis or recent surgical procedure
- Serious, non-healing or dehiscing wound, active ulcer, or untreated bone fracture
- Metastatic disease that involves major airways or blood vessels, or centrally located mediastinal tumor masses (< 30 mm from the carina) of large volume

  Patients with vascular invasion of the portal or hepatic veins may be enrolled.
- History of intra-abdominal inflammatory process within 6 months prior to initiation of study treatment, including but not limited to active peptic ulcer disease, diverticulitis, or colitis
- Radiotherapy within 28 days and abdominal/ pelvic radiotherapy within 60 days prior to initiation of study treatment, except palliative radiotherapy to bone lesions within 7 days prior to initiation of study treatment
- Local therapy to liver (e.g., radiofrequency ablation, percutaneous ethanol or acetic acid injection, cryoablation, high-intensity focused ultrasound, transarterial chemoembolization, transarterial embolization, etc.) within 28 days prior to initiation of study treatment or non-recovery from side effects of any such procedure
- Major surgical procedure, open biopsy, or significant traumatic injury within 28 days prior to initiation of study treatment, or abdominal surgery, abdominal interventions or significant abdominal traumatic injury within 60 days prior to initiation of study treatment or anticipation of need for major surgical procedure during the course of the study or non-recovery from side effects of any such procedure
- Chronic daily treatment with a non-steroidal anti-inflammatory drug (NSAID)

  Occasional use of NSAIDs for the symptomatic relief of medical conditions such as headache or fever is allowed.

End of Study

The end of this study is defined as the date when the last patient, last visit (LPLV) occurs (i.e., last patient in the global and extended China enrollment phases combined) or safety follow-up is received from the last patient (global and extended China enrollment phases combined), whichever occurs later.

In addition, the Sponsor may decide to terminate the study at any time.

Investigational Medicinal Products

Patients will receive treatment as outlined below until unacceptable toxicity or loss of clinical benefit as determined by the investigator after an integrated assessment of radiographic and biochemical data, and clinical status (e.g., symptomatic deterioration such as pain secondary to disease).

Arm	Cycle Length	Dose, Route, and Regimen (drugs listed in order of administration)	Infusion Rate
A [a]	21 days	Atezolizumab 1200 mg IV on Day 1	Over 60 (±15) minutes (for the first infusion); 30 (±10) minutes for subsequent infusions if tolerated
		Bevacizumab 15 mg/kg IV on Day 1	Over 90 (±15) minutes (for the first infusion); shortening to 60 (±10) then 30 (±10) minutes for subsequent infusions if tolerated
B	21 days	Sorafenib 400 mg BID, by mouth, continuously	Not applicable

BID = twice per day.

[a] For patients randomized to Arm A, on Day 1 of each cycle, atezolizumab will be administered first, followed by bevacizumab, with a minimum of 5 minutes between dosing.

Statistical Methods

Primary Analysis

The primary efficacy objective for this study is to evaluate the efficacy of atezolizumab in combination with bevacizumab compared with sorafenib on the basis of the co-primary efficacy endpoints of OS and IRF-assessed PFS according to RECIST v1.1.

OS and PFS will be tested $initially$ in parallel with the overall type I error controlled at a two-sided significance level of 0.05, where OS will be tested at a two-sided significance level of 0.048 and PFS will be tested at a two-sided significance level of 0.002.

Overall Survival

OS is defined as the time from the date of randomization to the date of death from any cause. Patients who are alive at the time of the analysis data cutoff will be censored at the last date they were known to be alive. Patients with no post-baseline information will be censored at the date of randomization.

The two-sided log-rank test, stratified by geographic region (Asia excluding Japan vs. rest of world), macrovascular invasion and/or extrahepatic spread (presence vs. absence), and baseline AFP ($<$400 vs. $\geq$400 ng/mL), will be used as the primary analysis to compare OS between the two treatment arms. The results from the unstratified log-rank test will also be provided as a sensitivity analysis to check the robustness of the results of the stratified log-rank test.

The Kaplan-Meier method will be used to estimate median OS for each treatment arm. Brookmeyer-Crowley methodology will be used to calculate the 95% CI for the median OS for each treatment arm. Stratified Cox proportional-hazards models will be used to estimate the HR and its 95% CIs. The stratification factors will be the same as those used for the primary stratified log-rank test. The unstratified HR will also be provided.

The final analysis of OS will occur after approximately 312 deaths have occurred.

A group sequential design will be used for testing the co-primary efficacy endpoint OS to account for the conduct of two interim analyses. An α spending using the Lan-De Mets method approximating O'Brien-Fleming boundaries will be utilized to control the overall Type I error rate of 0.048 for the OS co-primary efficacy endpoint. The $first$ interim OS analysis will be

conducted at time of the co-primary analysis of IRF-assessed PFS. *The second interim OS analysis will be conducted when approximately 243 OS events have been observed.*

IRF-Assessed Progression-Free Survival

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

Methods for PFS analyses are similar to those described for the OS endpoint.

Determination of Sample Size

A total of approximately 480 patients will be randomized in the global enrollment phase of this study, using a 2:1 randomization ratio to allocate patients to either the atezolizumab + bevacizumab arm (Arm A) or the sorafenib arm (Arm B). The final *OS* analysis will be conducted at approximately *33* months after the first patient is randomized (first patient in [FPI]).

The co-primary efficacy endpoints for this study are as follows:

- OS, defined as the time from randomization to death from any cause
- IRF-assessed PFS, defined as the time from randomization to the occurrence of disease progression per RECIST v1.1 or death from any cause, whichever occurs first.

The overall Type I error rate for this study will be controlled at a two-sided significance level of 0.05 by a *graphical approach (i.e., alpha splitting and recycling). The* overall two-sided significance level of 0.05 *will be* split into a two-sided significance level of 0.048 for the testing of OS and a two-sided significance level of 0.002 for the testing of PFS as a first step.

The sample size of the study was determined based on the number of deaths required to demonstrate efficacy in terms of OS. To detect an improvement in OS using a log-rank test at a two-sided significance level of 0.048, approximately *312* deaths will be required to achieve 80% *overall* power assuming a target HR of 0.71 (median OS improvement vs. control is 4.9 months). The minimum detectable difference (MDD) of OS is an HR of 0.*783* (median OS improvement is 3.3 months). This analysis is expected to occur approximately *33* months after FPI.

The calculation of sample size and estimates of the OS analysis timelines are based on the following assumptions:

- Patients will be randomized to the atezolizumab + bevacizumab and sorafenib arms in a 2:1 ratio.
- OS follows a one-piece exponential distribution.
- The median OS in the control arm is 12 months.
- The stopping boundaries of the interim and final analyses of OS use the O'Brien-Fleming boundaries approximated using the Lan-DeMets method.
- The dropout rate is 5% for the atezolizumab + bevacizumab arm and 10% for the sorafenib arm over 12 months for OS.
- The recruitment of approximately 480 patients will take place over approximately *10* months

Interim Analysis

No interim analysis is planned for PFS in this study.

Two interim analyses of OS will be performed. The first interim analysis will be performed at the time of the primary PFS analysis, estimated to occur at approximately *16* months after FPI. It is anticipated that at this time approximately *172* deaths will have been observed. The respective MDD OS HR is 0.*633* (median OS improvement is *6.9* months). *The second OS interim analysis is planned to be conducted when approximately 243 deaths are accumulated,*

estimated to occur at approximately 24 months after FPI. The respective MDD OS HR is 0.728 (median OS improvement is 4.6 months).

To control the two-sided significance level at 0.048 for the interim and final OS analyses, the Lan-DeMets method will be used to approximate the O'Brien-Fleming boundaries.

The planned interim analyses for OS will be conducted by the Sponsor.

Analysis Timing	Planned Information Fraction	Required Events/ Analysis timing *(estimated)*	Stopping Boundary (Two-Sided p-Value)	
			Alpha can be recycled to OS (i.e., OS alpha = 0.05)	*Alpha cannot be recycled to OS (i.e., OS alpha = 0.048)*
1st OS interim analysis	55%	*172/16* months	MDD.HR ≤ 0.*636* (p-value ≤ 0.005)	MDD.HR ≤ 0.*633* (p-value ≤ 0.*005*)
2nd OS interim analysis	*78%*	*243/24 months*	*MDD.HR ≤ 0.73 (p-value ≤ 0.021)*	*MDD.HR ≤ 0.728 (p-value ≤ 0.02)*
OS final analysis	100%	*312/33* months	MDD.HR ≤ 0.*784* (p-value ≤ 0.*043*)	MDD.HR ≤ 0.*783* (p-value ≤ 0.*041*)

HR = hazard ratio; MDD = minimum detectable difference; OS = overall survival; PFS=progression-free survival.

Analysis timing shown in the table is projected based on protocol assumptions. Actual timing depends on the exact time that the required events have accrued.

The 1st OS interim analysis will be conducted when approximately 308 PFS events have happened.

Appendix 2 Schedule of Activities

Assessment Window (Days) [a]	Screening [b]		Treatment Phase (Q3W)	Treatment Discontinuation [c]	Survival Follow-Up
	–28 to –1	–7 to –1	Day 1 of Each Cycle [c]	≤ 30 Days after Last Dose	
Signed Informed Consent Form(s) [b]	x				
Review of eligibility criteria	x				
Medical, surgical, and cancer histories, including demographic information [d]	x				
Complete physical examination [e]	x				
Limited physical examination [f]			x [g]	x	
ECOG Performance Status		x	x [g]	x	
Patient-reported outcomes [h]			x [g]	x	x [h]
Tumor assessment [i]	x		See footnote [i]	x	x
Vital signs [j]	x		x	x	
Weight	x		x [k]	x	
Height	x				
12-lead ECG [l]	x		Perform as clinically indicated		
EGD [m]	x				
Hematology [n, z]		x	x [g]	x	
Serum chemistry [o, z]		x	x [g]	x	
HIV, HBV, HCV serology [p]	x				

Appendix 2 Schedule of Activities (cont.)

Assessment Window (Days) [a]	Screening [b]		Treatment Phase (Q3W)	Treatment Discontinuation [c]	Survival Follow-Up
	–28 to –1	–7 to –1	Day 1 of Each Cycle [c]	≤ 30 Days after Last Dose	
Quantitative HBsAg, HBV DNA, HCV RNA [q]	x		x	x	
α-fetoprotein	x		x	x	
Coagulation panel (aPTT, INR) [z]		x	x [g]	x	
Urinalysis [r, z]		x	x [g]	x	
TSH, free T3, free T4	x		Cycles 5, 9, 13, etc. (every 4 cycles)	x	
Pregnancy test		x [s]	x [t]	x	
Serum PK sample			See Appendix 3		
Serum ADA sample			See Appendix 3		
Pharmacodynamic samples for biomarkers			See Appendix 3	x	
Plasma sample for RBR (optional)			Any time (at investigator's discretion)		
Archival tumor tissue sample (or optional fresh biopsy if archival tissue is not available) for biomarkers	x				
Concomitant medications [u]		x	x	x	
Adverse events [v]	x		x	x	
Study treatment infusion [w]			x		
Sorafenib dispensing [x]			x		
Survival and anti-cancer therapy follow-up [y]					x

Appendix 2 Schedule of Activities (cont.)

ADA = anti-drug antibody; CT = computed tomography; EGD = esophagogastroduodenscopy; EORTC = European Organisation for Research and Treatment of Cancer; EQ-5D-5L = EuroQol 5-Dimension Questionnaire, 5-level version; HBcAb = hepatitis B core antibody; HBsAb = hepatitis B surface antibody; HBsAg = hepatitis B surface antigen; HBV = hepatitis B virus; HCV = hepatitis C virus; imRECIST = immune-modified RECIST; MRI = magnetic resonance imaging; PET = positron emission tomography; PK = pharmacokinetic; PRO = patient-reported outcome; Q3W = every 3 weeks; QLQ-C30 = quality-of-life questionnaire for cancer; QLQ-HCC18 = HCC disease-specific module; RBR = Research Biosample Repository; RECIST v1.1 = Response Evaluation Criteria in Solid Tumors, Version 1.1; T3 = triiodothyronine; T4 = thyroxine; TSH = thyroid-stimulating hormone.

Note: Assessments scheduled on the days of study treatment infusions should be performed before the infusion unless otherwise noted. Each cycle is 21 days in length.

[a] The first dosing date (Cycle 1, Day 1) should occur within 3 business days from randomization, with the exception of the emergence of an adverse event for which dosing may be postponed. All visits and infusions thereafter may be administered with a window of ± 3 days.

[b] Written informed consent can be obtained up to 30 days prior to study entry and is required before performing any study-specific tests or procedures. Results of standard-of-care tests or examinations performed prior to obtaining informed consent and per protocol relevant window may be used for screening assessments rather than repeating such tests. Screening local laboratory assessments obtained ≤ 96 hours prior to the initiation of study treatment do not have to be repeated for Cycle 1. Test results should be reviewed prior to administration of study treatment.

[c] Patients will be asked to return to the clinic 30 days after the last dose of study treatment for an end-of-treatment visit. After this visit, serious adverse events and protocol defined adverse events of special interest, regardless of attribution, will be recorded until 90 days after the last dose of study treatment or until initiation of another anti-cancer therapy, whichever occurs first. Ongoing adverse events thought to be related to study treatment will be followed until the event has resolved to baseline grade or better, the event is assessed by the investigator as stable, new anti-cancer treatment is initiated, the patient is lost to follow-up, the patient withdraws consent, or it is determined that the study treatment or participation is not the cause of the adverse event. Scans performed within 6 weeks prior to the treatment discontinuation visit do not need to be repeated.

[d] Cancer history includes stage, date of diagnosis, and prior anti-tumor treatment. Demographic information includes age and self-reported race/ethnicity. Reproductive status and smoking history should also be captured.

[e] A complete physical examination at screening should include the evaluation of head, eye, ear, nose, and throat and cardiovascular, dermatologic, musculoskeletal, respiratory, gastrointestinal, genitourinary, and neurologic systems. Changes in abnormalities noted at baseline should be recorded at the end of the visit. New or worsened abnormalities should be recorded as adverse events if appropriate.

[f] A limited physical examination will be performed at other visits to assess changes from baseline abnormalities and any new abnormalities and to evaluate patient reported symptoms. New or worsened abnormalities should be recorded as adverse events if appropriate.

Appendix 2 Schedule of Activities (cont.)

[g] ECOG Performance Status, limited physical examination, local laboratory assessments*, and PROs* may be obtained ≤96 hours before Day 1 of each cycle.

[h] The EORTC QLQ-C30, EORTC QLQ-HCC18, and EQ-5D-5L questionnaires will be completed by all patients on paper starting on Day 1 of Cycle 1 and Day 1 of every cycle thereafter. All PRO questionnaires scheduled for administration during a clinic visit are required to be completed by the patient at the investigational site at the start of the clinic visit before discussion of the patient's health state, lab results or health record, before administration of study treatment, and/or prior to any other study assessment(s). *This is to avoid any potential bias to* patients' responses to ensure that the validity of the instrument is not compromised and that data quality meets regulatory requirements. *Interview assessment by a member of the clinic staff will be allowed if the patient is not able to complete the measure on their own.* Study personnel should review all questionnaires for completeness before the patient leaves the investigational site. During survival follow-up, all PRO questionnaires will be completed every 3 months (for 1 year), unless the patient withdraws consent or the Sponsor terminates the study. PRO questionnaires during the survival follow-up period may be completed at the investigational site should the patient come in for a clinic visit or be administered via *interview in* telephone calls.

[i] All measurable and evaluable lesions should be assessed and documented at the screening visit. Radiologic imaging performed during the screening period should consist of 1) CT and/or MRI of the chest/abdomen/pelvis and brain, 2) bone scan or PET scan as clinically indicated, and 3) any other imaging studies (CT scan of the neck, plain films, etc.) as clinically indicated by the treating physician. The same radiographic procedures and technique must be used throughout the study for each patient (e.g., if the patient had CT chest/abdomen/pelvis performed during screening, then she should subsequently undergo CT performed using the same radiologic protocol throughout the remainder of the study). Results must be reviewed by the investigator before dosing at the next cycle. Tumor assessments will be performed at baseline, every 6 weeks (±1 week) for the first 54 weeks following the initiation of study treatment, and every 9 weeks (±1 week) thereafter, with additional scans as clinically indicated. All known sites of disease documented at screening should be re-assessed at each subsequent tumor evaluation. Tumor response will be evaluated by the investigator using RECIST Version 1.1 and imRECIST. In the absence of disease progression, tumor assessments should continue regardless of whether patients discontinue study treatment or start new anti-cancer treatment, unless the patient dies, withdraws consent, or the study is terminated by the Sponsor, whichever occurs first.

[j] Vital signs include heart rate, respiratory rate, blood pressure, and temperature. For patients randomized to Arm A, on days of study treatment administration (atezolizumab and bevacizumab), the patient's vital signs should be determined up to 60 minutes before all infusions. Vital signs will be measured at the end of bevacizumab infusion and 2 (±1) hours after end of the infusion and will also be collected during and after every infusion of atezolizumab if clinically indicated.

[k] The dose of bevacizumab will be based on the patient's weight (in kilograms) measured ≤14 days prior to baseline (the initiation of study treatment) and will remain the same throughout the study unless there is a weight change of >10% from baseline. If re-baseline is needed the latest baseline weight should always be used to calculate percent change in weight for all subsequent doses.

[l] Patients should be resting and in a supine position for at least 10 minutes prior to each ECG collection.

Appendix 2 Schedule of Activities (cont.)

[m] All patients must undergo an EGD and all size of varices (small to large) must be assessed and treated per local standard of care prior to enrollment.

[n] Hematology consists of CBC, including RBC count, hemoglobin, hematocrit, WBC count with differential (neutrophils, eosinophils, lymphocytes, monocytes, basophils, and other cells), and platelet count. A manual differential can be done if clinically indicated.

[o] Serum chemistry includes bicarbonate or total carbon dioxide (if considered standard of care for the region), sodium, potassium, magnesium, chloride, glucose, BUN or urea, creatinine, total protein, albumin, phosphorus, calcium, total bilirubin, alkaline phosphatase, ALT, AST, and LDH.

[p] All patients will be tested for HIV locally prior to the inclusion into the study and if not in contradiction with local legislation; HIV-positive patients will be excluded from the clinical study. HBsAg, HBcAb, and HBsAb should be collected during screening and tested locally. HBV DNA must be collected prior to Cycle 1, Day 1 in patients who have negative serology for HBsAg and positive serology for anti HBcAb.

[q] Only if patient tests positive for HBsAg, HBcAb, quantitative HBsAg and HBV DNA will be tested during screening; Cycle 5, Day 1; Cycle 9, Day 1; and at treatment discontinuation. Quantitative HBsAg will be tested by central laboratory. If a patient tests positive for HCV antibody at screening, quantitative HCV RNA must be tested locally at screening, Cycle 5 Day 1, Cycle 9 Day 1, and at treatment discontinuation.

[r] Urine dipstick includes specific gravity, pH, glucose, protein, ketones, and blood and should be repeated before every cycle during treatment. Urine dipstick for proteinuria must be $<$2+ within 7 days prior to initiation of study treatment. Patients discovered to have $\geq$2+ proteinuria on dipstick urinalysis at baseline should undergo a 24-hour urine collection and must demonstrate $<$1 g of protein in 24 hours.

[s] Serum pregnancy test within 14 days before Cycle 1, Day 1.

[t] Urine pregnancy test; if a urine pregnancy test is positive, it must be confirmed by a serum pregnancy test.

[u] Concomitant medications include any prescription medications or over-the-counter medications. At screening, any medications the patient has used within the 7 days prior to initiation of study treatment should be documented. At subsequent visits, changes to current medications or medications used since the last documentation of medications will be recorded.

[v] After informed consent has been obtained but prior to initiation of study drug, only serious adverse events caused by a protocol-mandated intervention should be reported. After initiation of study drug, all adverse events will be reported until 30 days after the last dose of study treatment or until initiation of another anti-cancer therapy, whichever occurs first. Serious adverse events and adverse events of special interest will continue to be reported until 90 days after the last dose of study treatment or until initiation of new anti-cancer therapy, whichever occurs first. After this period, investigators should report any serious adverse events and adverse events of special interest that are believed to be related to prior treatment with study drug. The investigator should follow each adverse event until the event has resolved to baseline grade or better, the event is assessed as stable by the investigator, new anti-cancer treatment is initiated, the patient is lost to follow-up, the patient withdraws consent, or it is determined that the study treatment or participation is not the cause of the adverse event. Every effort should be made to follow all serious adverse events considered to be related to study drug or study-related procedures until a final outcome can be reported.

Appendix 2 Schedule of Activities (cont.)

[w] The initial dose of atezolizumab will be delivered over 60 (±15) minutes. If the first infusion is tolerated without infusion-associated adverse events, the second infusion may be delivered over 30 (±10) minutes. If the 30-minute infusion is well tolerated, all subsequent infusions may be delivered over 30 (±10) minutes. The initial dose of bevacizumab will be delivered over 90 (±15) minutes. If the first infusion is tolerated without infusion-associated adverse events, the second infusion may be delivered over 60 (±10) minutes. If the 60-minute infusion is well tolerated, all subsequent infusions may be delivered over 30 (±10) minutes. For patients randomized to Arm A, atezolizumab will be administered first followed by bevacizumab, with a minimum of 5 minutes between dosing. In the absence of unacceptable toxicity, patients may continue study treatment until there is evidence of disease progression or lack of clinical benefit.

[x] Sorafenib is taken by mouth twice a day continuously.

[y] Survival follow-up information will be collected via telephone calls, patient medical records, and/or clinic visits approximately every 3 months (±21 days) until death, loss to follow-up, or until study termination by the Sponsor. All patients will be followed for survival and new anti-cancer therapy (including targeted therapy and immunotherapy) information unless the patient requests to be withdrawn from follow-up; this request must be documented in the source documents and signed by the investigator. If the patient withdraws from study, the study staff may use a public information source (e.g., county records) to obtain information about survival status only.

[z] Local laboratory assessments from each cycle must be reviewed prior to study treatment administration for each cycle.

Appendix 3

Schedule of Biomarker, Pharmacokinetic, and Immunogenicity Samples

Visit	Timepoint	Sample Type
Day 1 of Cycle 1	Prior to any drug administration	• Atezolizumab PK (serum) [a] • Atezolizumab ADA (serum) [a] • Biomarker (plasma, serum)[b]
	30 min after end of atezolizumab infusion	• Atezolizumab PK (serum) [a]
Day 1 of Cycle 2	Prior to any drug administration	• Atezolizumab PK (serum) [a] • Atezolizumab ADA (serum) [a] • Biomarker (plasma, serum) [b]
Day 1 of Cycle 3, Cycle 8, Cycle 12, Cycle 16	Prior to any drug administration	• Atezolizumab PK (serum) [a] • Atezolizumab ADA (serum) [a]
Day 1 of Cycle 4	Prior to any drug administration	• Atezolizumab PK (serum) [a] • Atezolizumab ADA (serum) [a] • Biomarker (plasma, serum) [b]
Treatment discontinuation visit (≤30 days after last dose)	At visit	• Atezolizumab PK (serum) [a] • Atezolizumab ADA (serum) [a] • Biomarker (plasma, serum)

[a] For Arm A only.

[b] Biomarker samples must be collected for both arms before dosing. For Arm B: If the patient takes the morning dose before coming to the visit, samples should be taken before the evening dose.