BeiGene

STATISTICAL ANALYSIS PLAN

Study Protocol Number:	BGB-A317-301
Study Protocol Title:	A Randomized, Open-label, Multi-center Phase 3 Study to Compare the Efficacy and Safety of BGB-A317 versus Sorafenib as First-Line Treatment in Patients with Unresectable Hepatocellular Carcinoma
Date:	April 21, 2022
Version:	Final Version 2.0

SIGNATURE PAGE

TABLE OF CONTENTS

DOCUMENT REVISION HISTORY

Version number	Finalization Date	Summary of Change
1.0	January 10, 2021	
2.0	April 21, 2021	Introduction section: remove the PK analysis Section 4.0: new added section for primary estimand (sub-section 4.1). Section 7.3.6: Update the definition for concomitant medication to align with definition update in TEAE. Section 7.4.1: Restructure the section to fit estimand framework Section 7.4.2: Add more clarification about the test statistics for secondary endpoints ORR and PFS. Section 7.4.3: add subgroup analysis for Race and PD-L1 expression Section 7.5.2: To streamline the logic of deriving TEAE, TEAE definition is updated Section 7.5.2.1: Clarify that TEAE overview will not include imAE; Overview of TEAE will be summarized by region in safety analysis set Section 7.5.2.2: add sentence to emphasize all imAE up to last dose of tislelizumab/placebo will be summarized, clarity that imAE will be summarized in tislelizumab + chemotherapy arm only

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Term
ADA	Antidrug antibody
AE	Adverse event
BCLC	Barcelona Clinic Liver Cancer
BID	Twice daily (Bis In Die)
BIRC	Blinded Independent Review Committee
BOR	Best overall response
CBR	Clinical benefit rate
COVID-19	Coronavirus disease of 2019
DCR	Disease control rate
DOR	Duration of response
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic case report form
EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30
EORTC QLQ-HCC 18	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Hepatocellular Carcinoma 18 Questions
EQ-5D	European Quality of Life 5-Dimensions
HBV	Hepatitis B virus
HCC	Hepatocellular carcinoma
HCV	Hepatitis C virus
HR	Hazard ratio

HRQoL	Health-related quality of life
IDMC	Independent Data Monitor Committee
ITT	Intent-To-Treat
IV	Intravenous
IWRS	Interactive Web Response System
LKADT	Last known alive date
MedDRA	Medical Dictionary for Regulatory Activities
NA	Not assessable
NCI-CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
NE	Not evaluable
ORR	Objective response rate
OS	Overall survival
PD	Progressive disease
PD-L1	Programmed cell death ligand-1
PFS	Progression-free survival
PK	Pharmacokinetic(s)
PO	Orally
PP	Per-Protocol
PR	Partial response
RECIST	Response Evaluation Criteria in Solid Tumors
RMST	Restricted Mean Survival Time
ROW	Rest of world
RPSFT	Rank Preserving Structural Failure Time

Q2W	Once every 2 weeks
Q3W	Once every 3 weeks
QTcF	QT interval Fridericia's correction formula
SD	Stable disease
SOC	System Organ Class
TEAE	Treatment-emergent adverse event
TTR	Time to Response
TTP	Time to progression

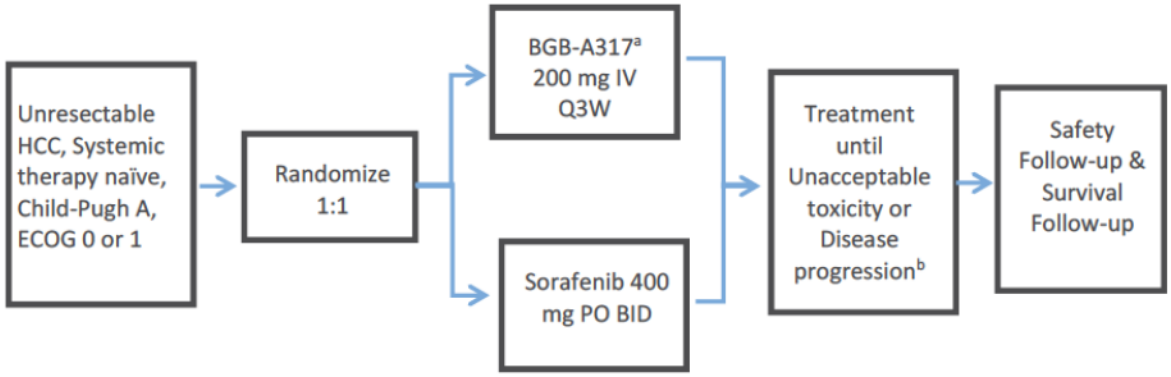
1 INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to describe the procedures and the statistical methods that will be used to analyze and report results for BGB-A317-301: A Randomized, Open-label, Multi-center Phase 3 Study to Compare the Efficacy and Safety of Tislelizumab versus Sorafenib as First-Line Treatment in Patients with Unresectable Hepatocellular Carcinoma (HCC). This SAP is based on BGB-A317-301 Protocol Amendment 5.0 dated 11 May 2020. The focus of this SAP is the planned interim and final analysis specified in the study protocol. The treatment information is masked to the statistician and study programming team throughout the study before final database lock.

2 STUDY OVERVIEW

2.1 STUDY DESIGN

This is a randomized, open-label, multicenter Phase 3 study to compare the efficacy and safety of tislelizumab to that of sorafenib as first-line treatment in adult patients with unresectable HCC. The study design schema is as follows:



a. The initial infusion (Cycle 1, Day 1) will be administered over a period of 60 minutes. If this infusion is well tolerated, subsequent infusions may be administered over 30 minutes. After tislelizumab infusion, patients will be further monitored for a period of 2 hours during Cycles 1 and 2. From Cycle 3 onward, a post-infusion monitoring period of ≥ 30 minutes will be required.
b. At the discretion of the Investigator, patients may be treated beyond disease progression under protocol-defined conditions (see Section 7.13.1 of the protocol).

Before initiating this Phase 3 study in Japan, a sub-study investigating the safety, tolerability, PK, and preliminary efficacy in Japanese patients is planned (see Appendix 13 of the protocol for details). A separate analysis plan is developed for Japan sub-study (Appendix 11.10).

After providing written informed consent, completing all Screening assessments, and being confirmed as eligible for study participation, approximately 640 patients will be randomized in a 1:1 ratio to receive (on an open-label basis) either tislelizumab or sorafenib.

At randomization, patients will be stratified by the following 5 factors:

- Macrovascular invasion (present vs absent)

- Extrahepatic spread (present vs absent)
- ECOG (0 vs 1)
- Etiology (hepatitis C virus [HCV] vs other [includes HBV])
- Geography (Asia [excluding Japan] vs Japan vs Rest of World)

Patients with HBV and HCV co-infection will be grouped along with HBV into the "other" category of etiology for randomization. Patients will then begin open-label treatment with one of the following regimens:

- Arm A: tislelizumab 200 mg intravenously (IV) once every 3 weeks (Q3W)
- Arm B: Sorafenib 400 mg orally (PO) twice daily (BID)

All study treatment is to be administered until intolerable toxicity, withdrawal of informed consent, or the time point at which, in the opinion of the Investigator, the patient is no longer benefiting from study therapy.

Treatment beyond the initial Investigator-assessed, RECIST v1.1-defined disease progression is permitted in both treatment arms provided the patient meets the criteria described in protocol Section 7.13.1.

2.2 Study Assessments

Tumor response will be evaluated by the BIRC and Investigator every 9 weeks during the first year of treatment and every 12 weeks from the second year onwards, in accordance with RECIST v1.1. If a patient discontinues study treatment due to any reason other than disease progression or death, tumor assessments will continue as scheduled until disease progression, death, loss to follow-up, withdrawal of consent, or until the study terminates, whichever occurs first.

Patients will be evaluated for any AEs and serious adverse events (SAEs) occurring up to 30 days after the last dose of study drug (all severity grades, per National Cancer Institute Common Terminology Criteria for Adverse Events [NCI CTCAE] Version [v] 4.03) or initiation of new anticancer therapy, whichever occurs first, and for all immune-related adverse events (irAEs) (tislelizumab arm only) occurring up to 90 days after the last dose of tislelizumab, regardless of whether the patients start a new anticancer therapy. All drug-related SAEs will be recorded by the Investigator after treatment discontinuation until patient death or loss to follow-up, whichever occurs first.

Safety and efficacy monitoring will be performed by an Independent Data Monitoring Committee (IDMC). The IDMC may recommend modifications to the study, including termination due to safety and/or efficacy concerns. The functions and membership of the IDMC are described in the IDMC Charter.

3 STUDY OBJECTIVES

3.1 Primary Objectives

- To compare overall survival (OS) between tislelizumab and sorafenib as first-line treatment in patients with unresectable hepatocellular carcinoma (HCC)

3.2 SECONDARY OBJECTIVES

- To compare objective response rate (ORR) assessed by Blinded Independent Review Committee (BIRC) according to Response Evaluation Criteria in Solid Tumors (RECIST) Version (v)1.1 between tislelizumab and sorafenib
- To compare progression-free survival (PFS) assessed by BIRC according to RECIST v1.1 between tislelizumab and sorafenib
- To compare duration of response (DOR) assessed by BIRC according to RECIST v1.1 between tislelizumab and sorafenib
- To compare time to progression (TTP) between tislelizumab and sorafenib
- To compare health-related quality of life (HRQoL) between tislelizumab and sorafenib
- To compare tumor assessment outcomes (i.e., ORR, PFS, DOR, TTP) assessed by the investigator according to RECIST v1.1 between tislelizumab and sorafenib
- To compare disease control rate (DCR) and clinical benefit rate (CBR), assessed by BIRC and investigator according to RECIST v1.1, between tislelizumab and sorafenib
- To compare safety and tolerability of tislelizumab versus sorafenib

3.3 EXPLORATORY OBJECTIVES

- To explore potential predictive biomarkers, including but not limited to programmed cell death ligand-1 (PD-L1) expression
- To characterize the pharmacokinetics (PK) of tislelizumab in patients with unresectable HCC
- To determine host immunogenicity to tislelizumab in patients with unresectable HCC

4 DEFINITION OF PRIMARY ESTIMAND

4.1 PRIMARY ESTIMAND – EQUIVALENT SURVIVAL BENEFIT AS SORAFENIB IN UNRESECTABLE HEPATOCELLULAR CARCINOMA

The primary analysis was written in estimand framework per study design as described in the protocol.

The primary scientific question of interest: will tislelizumab and sorafenib have equivalent survival benefit in first-line unresectable HCC, regardless of whether patients receive anticancer therapy subsequently.

The primary estimand is described by the following attributes:

1. Treatment of interest:

   The experimental treatment regimen is tislelizumab. The control treatment regimen is sorafenib.

2. Population:

   Adult patients with unresectable HCC, who have not received previous systemic therapy for advanced disease.

3. Primary variable:
   Overall survival defined as the time from the date of randomization to the date of death due to any cause. Further details on OS are provided in Section 7.4.1.
4. Handling of intercurrent events:
   - New anticancer therapy any incidence will be ignored, i.e., any death or patients' data collected after the new anticancer therapy will be considered for analysis (treatment policy strategy).
   - Discontinuation of treatment: any death or patients' data collected after the discontinuation of treatment will be considered for analysis (treatment policy strategy).
   - Death due to COVID-19 infection: will be counted as an event in the analysis of OS (composite strategy).
   - Treatment interruption/discontinuation due to COVID-19 infection: any incidence will be ignored. Any death or patients' data collected after the treatment interruption/discontinuation due to COVID-19 infection will be considered for analysis (treatment policy strategy).
   - Any other unforeseen intercurrent events: OS will take into account all deaths and any patients' data after any unforeseen intercurrent events.
5. Population-level summary:
   Hazard ratio (HR) of OS comparing tislelizumab versus sorafenib using Cox proportional hazard model stratified by the actual pooled stratification factors including region (Asia vs. EU/US), macrovascular invasion and/or extrahepatic spread (present vs. absent), etiology (HCV vs. other), and ECOG (0 vs. 1).

5 STUDY ENDPOINTS

5.1.1 Primary Endpoints

- OS – defined as the time from the date of randomization to the date of death due to any cause

5.1.2 Secondary Endpoints

- ORR as assessed by BIRC – defined as the proportion of patients with a confirmed CR or PR per RECIST v1.1
- PFS – defined as the time from the date of randomization to the date of the first objectively documented tumor progression, assessed by BIRC per RECIST v1.1, or death, whichever occurs first
- DOR – defined as the time from the first determination of an objective response, assessed by BIRC per RECIST v1.1, until the first documentation of progression or death, whichever occurs first
- TTP – defined as the time from the date of randomization to the date of the first objectively documented tumor progression, assessed by BIRC per RECIST v1.1

- HRQoL – measured using European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Hepatocellular Carcinoma 18 Questions (EORTC QLQ HCC 18) index score, European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30 (EORTC QLQ-C30) index-score, and the European Quality of Life 5 Dimensions, 5-level (EQ5D-5L) scale
- Tumor assessment (i.e., ORR, PFS, DOR and TTP), assessed by investigator per RECIST v1.1
- DCR – defined as the proportion of patients whose best overall response (BOR) is CR, PR, or SD, assessed by BIRC and investigator per RECIST v1.1
- CBR – defined as the proportion of patients who have CR, PR, or SD of ≥ 24 weeks in duration, assessed by BIRC and investigator per RECIST v1.1
- Safety assessment (e.g., new AEs, AEs present at baseline that worsen in severity during the study, and clinical laboratory abnormalities)

5.1.3 Exploratory Endpoints

- PD-L1 expression and other potential predictive biomarkers
- Summary of serum concentrations of tislelizumab
- Assessments of immunogenicity of tislelizumab by determining the incidence of antidrug antibodies (ADAs)

6 SAMPLE SIZE CONSIDERATIONS

The sample size calculation is based on the number of events required to demonstrate the noninferiority and superiority of tislelizumab over sorafenib (i.e., Arm A over Arm B) in the comparison of OS.

Under the original OS HR assumption of 0.75 (13.3 months in Arm A versus 10.0 months in Arm B), approximately 504 deaths in total were planned to have approximately 89% power in the superiority test and 98.4% power in the noninferiority test with a 1.08 noninferiority margin. A total of 640 patients were to be randomized in a 1:1 ratio to Arms A and B over a 16-month period at a constant enrollment rate (40 patients/month). An interim analysis was planned after 75% of the targeted number of OS events (i.e., approximately 378 deaths) occurred.

After reviewing recent published result (Yau et al. ESMO 2019, Abstract # 6572), a delayed effect of nivolumab vs sorafenib was observed which was manifested by the Kaplan-Meier curves of the two arms overlapping in the first 7-8 months before separating. The hazard ratio of 0.85 in CheckMate 459 also indicated a weaker than expected treatment effect possibly due to the confounding effect from subsequent immunotherapy use. Based on these observations, the interim analysis was postponed until approximately 80% of the targeted number of OS events (i.e., 403 deaths) were observed, with the planned number of deaths in the final analysis remaining at approximately 504. At the time of protocol amendment 5.0 (May 2020), enrollment was completed with a total of 674 patients randomized. Using a more conservative HR assumption of 0.79 at the time of final analysis after an initial 7-month delayed treatment effect (i.e., assuming

HR = 1 in the first 7 months), the estimated powers for the superiority test are 44% in the interim analysis and 72% in the final analysis using a simulation. The targeted number of events are estimated to occur approximately 33.9 and 46.6 months after study initiation under the actual enrollment rates, updated HR assumptions, and a median OS of 13.5 months in Arm B. The power in the noninferiority test is 0.935 under the updated assumptions.

In the comparison of the key secondary endpoint, ORR, between the 2 treatment arms, the power of a Miettinen and Nurminen test (Miettinen and Nurminen 1985) comparing 2 binomial rates for superiority in 674 patients is approximately 99%, assuming an ORR of 0.15 and 0.05 in Arms A and B, respectively.

7 STATISTICAL METHODS

7.1 ANALYSIS SETS

Intent-to-Treat (ITT) Analysis Set – Includes all randomized patients. Patients will be analysed according to their randomized treatment arm (i.e., either tislelizumab or sorafenib). This will be the primary analysis population for all efficacy analyses.

Per-Protocol (PP) Analysis Set – Includes all randomized patients who received at least one dose of their assigned study drug (tislelizumab or sorafenib) and had no critical protocol deviations (see Section 7.3.2 for details). Critical protocol deviations will be determined and documented before the database lock for the primary analysis.

Safety Analysis Set – Includes all randomized patients who received at least one dose of their assigned study drug (tislelizumab or sorafenib). This will be the analysis population for all safety analyses. Patients will be classified and analysed according to the study treatment they received Treatment received is defined as (i) the intended treatment if it was received at least once, or (ii) the first treatment received when starting therapy with study medication if intended treatment is never received. Each patient will be classified into and analysed consistently within one (and only one) treatment arm.

PK Analysis Set – included all patients who received at least one dose of tislelizumab per the protocol for whom any post dose PK data were available.

Antidrug antibody (ADA) Analysis Set – included all patients who received at least one dose of tislelizumab for whom both baseline ADA and at least one postbaseline ADA results were available.

7.2 DATA ANALYSIS GENERAL CONSIDERATIONS

7.2.1 Definitions and Computations

Study day: Study day will be calculated in reference to the first dose date. For assessments conducted on or after the date of first dose date, study day will be calculated as (assessment date –first dose date + 1). For assessments conducted before first dose date, study day is calculated as (assessment date – first dose date). There is no study day 0.

Baseline Measurements:

- For efficacy evaluation: a baseline value is defined as the last non-missing value

collected on or prior to randomization.

- For safety: a baseline value is defined as the last non-missing value on or prior to the first study drug administration.
- For toxicity grade of certain laboratory tests: two baseline toxicity grades should be derived according to the directions (lower (Hypo) or higher (Hyper)). For example, a baseline hemoglobin with value between 10.0 g/dL and Lower limit of normal, two baseline toxicity grades: Grade 1 for Hypo and Grade 0 for Hyper will be derived.

Study Follow-up Duration (SFD): Study follow-up duration is defined as the duration from the randomization date to the study discontinuation date (e.g. death, consent withdrawal, lost to follow-up) or to the cutoff date if a patient is still ongoing.

All calculations and analyses will be conducted using SAS version 9.2 or higher.

7.2.2 Conventions

Unless otherwise specified, the following conventions will be applied to all analyses:

- 1 year = 365.25 days. Number of years is calculated as (days/365.25) rounded up to 1 significant digit.
- 1 month = 30.4375 days. Number of months is calculated as (days/30.4375) rounded up to 1 significant digit.
- Age will be calculated as the integer part of (date of informed consent – date of birth + 1)/365.25.
- P-values will be rounded to 4 decimal places. P-values that round to 0.0000 will be presented as '< 0.0001' and p-values that round to 1.000 will be presented as '> 0.9999'.
- Time-to-event or duration of event endpoints based on tumor assessment will be based on the actual date the radiograph was obtained rather than the associated visit date.
- Missing efficacy or safety data will not be imputed unless otherwise specified.
- For laboratory results collected as < or >, a numeric value, 0.0000000001 will be subtracted or added, respectively, to the value.
- For by-visit observed data analyses, percentages will be calculated based on the number of patients with non-missing data as the denominator, unless otherwise specified.
- For continuous endpoints, summary statistics will include n, mean, standard deviation, median, Q1, Q3, and range (minimum and maximum).
- The unit of time duration is month(s) unless otherwise specified.

7.2.3 Handling of Missing Data

Missing data will not be imputed unless otherwise specified elsewhere in this document.

Specific rules for handling of missing or partially missing dates for adverse events, prior/concomitant medications/procedures, and subsequent anti-cancer therapy are provided in Appendix 11.1, 11.2, 11.3, and 11.4 .

By-visit endpoints will be analyzed using observed data, unless otherwise specified. For observed data analyses, missing data will not be imputed and only the observed records will be included.

7.2.4 Adjustments for Covariates

The value of the pooled stratification factors used at randomization (IWRS), including region (Asia; ROW), etiology, macrovascular invasion and/or extrahepatic spread (Yes; No), and ECOG, will be used in the stratified log-rank test and stratified Cox proportional hazard model for the primary endpoint OS, the key secondary endpoint ORR, and other secondary endpoints. The actual value of the stratification factors (collected in eCRF) and other baseline factors may be used in the model as covariates as supportive analyses for endpoints. Additional adjustment for covariate(s) might be used for sensitivity analyses if an unexpected imbalance of baseline characteristics are identified after data base lock.

7.2.5 Multiplicity Adjustment

The primary and key secondary efficacy endpoints will be tested using a gate-keeping approach. The hypothesis testing will be performed sequentially in the following order: noninferiority of OS, superiority of OS, superiority of ORR by BIRC, and superiority of PFS by BIRC. At interim, noninferiority of OS will be skipped since the study will not stop with noninferiority of OS declared at interim. The primary objective of OS will be considered as achieved if noninferiority of OS is demonstrated at final analysis.

There will be one interim analysis of OS superiority The O'Brien-Fleming alpha-spending function is applied to the OS superiority test to control the overall type I error of 0.025 in this group sequential trial. The stopping boundary at the interim and final analyses will be determined by the actual number of death events using the O'Brien-Fleming stopping boundary approximated by the Hwang-Shih-DeCani spending function by setting gamma parameter at -4.

Stopping boundaries (p-value and Z score) of the superiority test for OS at the interim and final analyses are shown in Table 1. The boundaries will be updated from the actual death events observed at the time of the interim analysis and final planned number of deaths, using the alpha-spending function.

Table 1 Stopping Boundaries and Approximate Hazard Ratio Threshold for Interim and Final Analyses of Superiority Test for Overall Survival

	Time (months)	# Deaths	p-value (Z score) for Efficacy	Approximate Hazard Ratio for Efficacy
Interim analysis	33.9	403	< 0.0110 (> 2.29)	< 0.7958
Final analysis	46.6	504	< 0.0221 (> 2.01)	< 0.8358

At the interim analysis, if the null hypothesis of OS superiority is rejected, ORR by BIRC will be tested at one-sided alpha level 0.025. If ORR is rejected, PFS by BIRC will be tested at one-sided alpha level 0.025.

At the final analysis, if the null hypothesis of OS superiority is rejected, the same ORR statistics

calculated at the interim analysis will be tested at one-sided alpha level 0.025. If ORR is rejected, the same PFS statistics calculated at the interim will be tested at one-sided alpha level 0.025. The key secondary objectives of ORR and PFS will be considered as achieved if statistics calculated at the interim are significant at one-side alpha level 0.025.

The type I error for OS non-inferiority at the interim analysis was estimated as 0.00108 which was the probability of crossing the interim superiority boundary (<0.7958) under the noninferiority null hypothesis HR >= 1.08. By using the spending function interpolated method, the nominal p-value at the final analysis for non-inferiority is 0.02498. The non-inferiority of OS comparison between two treatment arms is claimed if the upper limit of the 95.004% confidence interval of hazard ratio < 1.08.

7.2.6 Data Integrity

Before the pre-specified interim or final statistical analysis begins, the integrity of the data should be reviewed to ensure it is fit-for-purpose. The data set for analysis should be an accurate and complete representation of the patients' relevant outcomes from the clinical database. All data should be complete and reviewed up to a pre-specified cutoff date. Consistency checks and appropriate source data verification should be complete.

A data integrity protection plan has been implemented in this study. Data summary by treatment arm is not accessible to the Sponsor. The IDMC meetings for safety review and pre-specified IA for efficacy are handled externally by an Independent Data Monitoring Committee (IDMC.)

7.3 Patient Characteristics

7.3.1 Patient Disposition

The number (percentage) of patients who signed informed consent, enrolled in the study, died before enrollment, and who were screen-fail failures will be summarized. Reasons for screen failures will also be summarized.

The number (percentage) of patients randomized, treated, discontinued from study drug and discontinued from the study will be summarized. The primary reason for end of treatment (study drug discontinuation) and end of study will be summarized by categories. The reasons for treatment/study discontinuation related to COVID-19 impact will also be summarized. Study follow up duration will be summarized descriptively.

Patient disposition will also be summarized by region (Asia [excluding Japan], Japan, and EU/US) for the ITT analysis set.

7.3.2 Protocol Deviations

Protocol deviation criteria will be established together with its category/term of important and not important. Important protocol deviation is defined as an event related to study inclusion or exclusion criteria, conduct of the trial, patient management, patient assessment, and/or a significant non-compliance that significantly affects or has the potential to significantly affect human subject protection or reliability of the trial's results, per International Council for Harmonisation (ICH) Guideline E3. Patients with important protocol deviations will be identified and documented before the database lock. Important protocol deviations will be

summarized for all patients in the ITT analysis set. Critical protocol deviation that may significantly impact efficacy will be reviewed prior to data base lock according to the criteria defined in protocol deviation specification. Patients with the following critical protocol deviations will be excluded from per protocol analysis set:

- Did not have histologically confirmed diagnosis of HCC
- Did not sign the informed consent form at study entry
- Received different study drug from assigned treatment
- The ECOG PS score at screening was greater than 1
- Received any concurrent antineoplastic therapy during treatment period (i.e., chemotherapy, hormonal therapy, immunotherapy, standard/investigational agents [including Chinese (and other country) herbal medicine and patent medicines] for the treatment of cancer) (Appendix 11.9), or extensive radiation therapy (except for local, palliative radiotherapy to bone)

Protocol deviations that are related to COVID-19 will be summarized and all patients affected by COVID-19-related protocol deviations will be listed along with site information and specific deviation description.

7.3.3 Demographic and Other Baseline Characteristics

Demographic and other baseline characteristics will be summarized in the ITT analysis set using descriptive statistics. Continuous variables will be summarized using number of patients, mean, standard deviation, median, minimum, and maximum. Categorical variables will be summarized using number of patients and percentage in relevant categories.

Demographic and other baseline characteristics include:

- Age
- Age group (< 65 vs ≥ 65 years)
- Sex
- Race
- Ethnicity
- Geographic Region (Asia [excluding Japan], Japan, EU/US)
- Height (cm)
- Weight (kg)
- ECOG PS
- Tobacco Usage Status
- Alcohol Consumption Status

Demographic and other baseline characteristics will also be summarized by region (Asia [excluding Japan], Japan and EU/US) for ITT analysis set.

In addition, the stratification factors per IWRS and per eCRF will be summarized based on ITT analysis set:

- Macrovascular Invasion (yes vs. no)
- Extrahepatic Spread (yes vs. no)
- ECOG (0 vs 1)
- Etiology (hepatitis C virus [HCV] vs other [includes hepatitis B virus (HBV)])
- Geography (Asia [excluding Japan] vs Japan vs Rest of World)

7.3.4 Disease History and Baseline Disease Characteristics

The following disease history and baseline disease characteristics will be summarized in the ITT analysis set:

- Time since initial cancer diagnosis to date of randomization
- BCLC Initial Staging
- BCLC stage at study entry
- Child-Pugh (CP) classification/scores
- Histological differentiation grade
- Macrovascular invasion
- Extrahepatic spread
- Distant metastases and location
- Number of metastatic sites involved (0, 1, 2, ≥3)
- Hepatitis virus infection status (hepatitis B, hepatitis C, non-viral)
- Prior loco-regional procedures (present vs. absent)
- Alpha-fetoprotein at baseline (numeric, and categories: < 200 ug/L, >= 200 ug/L, < 400 ug/L, >=400 ug/L)

The hepatitis virus infection status comes from the relevant medical history collected from HCC history and characteristics case report form.

Detail of HCC relevant medical history including but not limited to Hepatitis B, Hepatitis C, History of Alcohol Abuse, baseline Child-Pugh score with its components, and HBV/HCV laboratory test values at baseline will also be summarized.

Cancer associated symptoms and symptoms with CTCAE grade 3 or 4 at baseline will also be summarized by SOC and preferred term e.

Disease history and baseline disease characteristics will also be summarized by region (Asia [excluding Japan], Japan, and EU/US) for ITT analysis set.

7.3.5 Prior Anti-Cancer Therapies and Surgeries

7.3.5.1 Prior Systemic Therapy

Patient data listings of prior systemic therapy will be provided.

7.3.5.2 Prior Anti-Cancer Radiotherapy [Except Liver]

The number (percentage) of patient with at least one prior anti-cancer radiotherapy [except liver], radiation therapy type, treatment intent, treatment setting, time from last radiotherapy to study entry (months), and treatment site (irradiated) will be summarized using the ITT analysis set.

7.3.5.3 Prior Anti-Cancer Surgeries/Procedures [Except Liver]

The number (percentage) of patients with at least one prior anti-cancer surgery, curative intent, time from end of last surgery to study entry (months), and anatomic location will be summarized using the ITT analysis set.

7.3.5.4 Prior Liver Local Regional Therapy

The number (percentage) of patients with at least one prior liver local regional therapy, therapy type, treatment intent, and site of local regional therapy will be summarized using the ITT analysis set.

Patient data listings of radio therapy [except liver], surgeries [except liver], and liver local regional therapy will be provided.

7.3.6 Prior and Concomitant Medication and Therapy

Prior and concomitant medications will be coded using the World Health Organization Drug Dictionary (WHO DD) drug codes and will be further classified to the appropriate Anatomical Therapeutic Chemical (ATC) code.

The number (percentage) of patients reporting prior and concomitant medications will be summarized by ATC medication class and WHO DD preferred term by phase in the safety analysis set. Prior medications are defined as medications that stopped before the first dose date. Concomitant medications will be defined as medications that (1) started before the first dose of study drug and were continuing at the time of the first dose of study drug, or (2) started on or after the date of the first dose of study drug up to 30 days after the patient's last dose or initiation of a new anti-cancer therapy, whichever occurs earlier. Patients who received concomitant systemic corticosteroid/immunosuppressant will also be summarized. Medication to treat COVID-19-related adverse events will also be summarized separately.

Patient data listings of prior and concomitant medication will be provided.

7.3.7 Medical History

Medical History will be coded the latest MedDRA version. The number (percentage) of patients reporting a history of any medical condition, as recorded on the eCRF, will be summarized by system organ class and preferred term in the safety analysis set.

A listing of medical history will be provided.

7.4 Efficacy Analysis

Overall survival is the primary endpoint of the study. One interim analysis of OS for efficacy is planned when approximately 80% (403) of the targeted death events (504) have been observed. Only superiority of OS will be tested at the IA and the noninferiority test will be skipped. At the final analysis, the superiority of tislelizumab over sorafenib will be tested for OS using a stratified log-rank test in the ITT analysis set only when noninferiority is demonstrated.

7.4.1 Primary Efficacy Endpoints

The primary estimand is defined in Section 4. Details of the statistical methods used in OS derivation and analysis are provided in this section including pre-defined sensitive analyses of the primary estimand and supplementary analyses of OS.

Variable:

Overall survival is defined as the time from randomization date to the documented death date for patients who died prior to or on the clinical cutoff date. For patients who are alive by the clinical cutoff date, OS will be censored at the last known alive date (LKADT). The LKADT will be defined as either the clinical data cutoff date for patients who are still on treatment, or last known alive date or cut-off date whichever comes first for other patients who are alive.

Every effort should be made to ensure collection of the complete date of death. In the rare case where the day of death is missing, the death date is imputed as the maximum (last available date showing patients alive + 1, first day of month of death date). The patient with imputed death date will be considered as an event for OS analysis. Death with missing month and/or year will not be imputed for OS analysis and will not be considered as OS event and will be censored at LKADT.

Primary efficacy analysis

OS will be compared between tislelizumab (Arm A) and sorafenib (Arm B) via testing the null hypothesis of noninferiority:

H_0: $HR_{A/B} \geq 1.08$

H_a: $HR_{A/B} < 1.08$

where $HR_{A/B}$ denotes the true hazard ratio for tislelizumab versus sorafenib.

Noninferiority of Overall Survival Margin Justification:

Data from 2 sorafenib trials (SHARP and Asia-Pacific) are used to derive the noninferiority margin. In the SHARP trial, the estimated HR (sorafenib/placebo) is 0.69 with a 95% CI of (0.55, 0.87). In the Asia-Pacific trial, the estimated HR (sorafenib/placebo) is 0.68 with a 95% CI of (0.50, 0.93). The pooled HR and its 95% CI is estimated as 0.6865 (0.5709, 0.8255) according to meta-analysis method (Parmar 1998). Using the 95% CI lower limit method on log HR (Rothmann 2003), the noninferiority margin corresponding to 60% retention of sorafenib effect over placebo is calculated as 1.08.

The 95.004% CI of $HR_{A/B}$ will be estimated in the ITT analysis set using a Cox proportional hazard model with treatment arm as a factor and stratified by the actual pooled stratification (i.e., collected

via eCRF) factors including region (Asia vs. EU/US), macrovascular invasion and/or extrahepatic spread (present vs. absent), etiology (HCV vs. other), and ECOG (0 vs. 1). Efron's method will be used in tie handling. Noninferiority will be declared if the upper limit of the 95.004% CI for $HR_{A/B}$ is less than 1.08.

At the final analysis, the superiority of tislelizumab over sorafenib will be tested for OS using a stratified log-rank test in the ITT analysis set only when noninferiority is demonstrated. Superiority will be declared if the one-sided p-value crosses the boundary specified in Section 7.2.5 in favor of Arm A in the stratified log-rank test at the interim or at the final analysis.

The median OS and the cumulative probability of OS estimated every 3 months and then 24 months will be calculated using Kaplan-Meier estimates for each treatment arm and presented with the 2-sided 95% CIs computed by the Brookmeyer and Crowley method using log-log transformation. The following sensitivity analyses and supportive analysis are planned:

Sensitivity analysis 1: An analysis of OS will be repeated based on the PP analysis set

Sensitivity analysis 2: An unstratified analysis of OS in the ITT analysis set

Sensitivity analysis 3: An analysis of OS based on stratification factors value collected at randomization (i.e., collected via IWRS).

Supportive analysis 4: An analysis to assess the proportional hazard assumption including Schoenfeld residual plot and time dependent covariate in the Cox model

The following supplementary analyses are planned:

Supplementary analysis 1: OS analysis based on the Max-Combo method will be performed to account for the possible non-proportional hazard effects. The weighted hazard ratio of OS (combo of G(0,0), G(0,1), G(1,0), and G(1,1)) from the Max-combo test will be displayed.

Supplementary analysis 2: OS analysis based on Restricted Mean Survival Time method (RMST) (RMST, Uno H, Claggett B, Tian L, Inoue E, et al. 2014). In order to account for the possible non-proportional hazard effect, the RMST will be computed for OS separately using the area under the curve from baseline to the minimum of the largest observed time on each of the two treatment groups.

Supplementary analysis 3: this analysis will aim at assessing the treatment effect based on OS had the COVID-19 pandemic not occurred. The primary variable is defined as the time from the date of randomization to the date of death due to non-COVID-19 pandemic reasons. The remaining intercurrent events will be handled as follows:

- **Discontinuation of study treatment due to any non-COVID-19 pandemic reasons**: OS will take into account all deaths irrespective of the study treatment discontinuation reasons (treatment policy strategy).
- **Discontinuation of study treatment due to COVID-19 pandemic reasons**: OS will be censored on the date of discontinuation of treatment due to the COVID-19 pandemic. The discontinuation reason due to the COVID-19 pandemic will be identified from the eCRF (hypothetical strategy).

- **Medications used for treating suspected or confirmed COVID-19 cases:** OS will be censored on the date of administration of COVID-19 medication
- **Death due to COVID-19**: OS will be censored on the date of death due to COVID-19.

Additional supplementary analyses may be carried out to explore the impact of next-line anticancer therapy for OS evaluation.

7.4.2 Secondary Efficacy Endpoints

7.4.2.1 Objective Response Rate by BIRC

The key secondary endpoint, confirmed objective response rate by BIRC per RECIST v1.1, defined in section 5.1.2, will be compared between tislelizumab and sorafenib via testing the null hypothesis:

H_0: ORR in Arm A = ORR in Arm B

H_a: ORR in Arm A > ORR in Arm B

Once superiority of OS is demonstrated, the statistical significance of the difference in ORR between arms in the ITT analysis set based on statistics observed in interim analysis will be evaluated using the Cochran-Mantel-Haenszel chi-square test with the actual value of the selected stratification factors as strata, tested at an alpha level of 0.05 (2-sided). Refer to section 7.2.5 for more details on multiplicity adjustment. The 2-sided 95% CIs for the odds ratio and the difference in ORR will be calculated, as well as Clopper-Pearson 95% CIs for the ORR within each arm.

Best overall response (BOR) was defined as the best response recorded from randomization until PD or the start of new anticancer treatment. Patients with no post-baseline response assessment (due to any reason) will be considered not assessable (NA) for BOR. The proportion of each of the response categories (CR, PR, SD, PD, NE, and NA) will be presented by treatment arm.

7.4.2.2 Progression-Free Survival by BIRC

PFS is defined as the time from the randomization date to disease progression or death, whichever occurs first. PFS will be censored at the last adequate tumor assessment if one of the following occurs: absence of event; the event occurred after a new anticancer therapy is given; the event occurred after two or more consecutive missing tumor assessments(Appendix 11.5).

The median PFS and the cumulative probability of PFS estimated at every 3 months will be calculated using Kaplan-Meier estimates for each treatment arm and presented with 2-sided 95% CIs computed by Brookmeyer and Crowley method using the log-log transformation. Once ORR is significant, the PFS in the ITT analysis set based on statistics observed in interim analysis will be evaluated using a stratified log-rank with the actual value of the selected stratification factors as strata.

The concordance/discordance of PFS by BIRC and investigators will be assessed by comparing PFS event type/censor between IRC and investigators, as well as summary of gap time in months for PFS as per IRC and as per investigators.

The following sensitivity analyses are planned:

- To evaluate the robustness of PFS per RECIST 1.1, two analyses of PFS by BIRC with different set of censoring rules will be performed:
    - The first sensitivity analysis is the same as the primary analysis except that it used the actual reported date of progression or death to define PFS regardless of whether new anticancer therapy is given
    - The second sensitivity analysis is the same as the primary analysis except that it used the actual reported date of progression or death to define PFS regardless of missing tumor scans.
- An unstratified analysis of PFS by BIRC in the ITT analysis set

7.4.2.3 Duration of Response, Time to Response, and Time to Progression by BIRC

Duration of Response (DOR) is defined in section 5.1.2 as progression/death event free time counted from the first objective response date to the first documented radiological PD date/or death date, whichever occurred first. All the censoring rules for PFS should be applied to DOR too. Time to Response (TTR) is defined as time from randomization to first objective response date.

Time to Progression (TTP) is defined as the time from the date of randomization to the date of the first objectively documented tumor progression, assessed per RECIST v1.1. TTP censoring rules are describe in Table .

Table 2 Censoring Rules for TTP

No	Situation	Date of Progression or Censoring	Primary Analysis
1	No baseline or post baseline tumor assessments	Randomization date	Censored
2	Progression documented between scheduled visits	Date of first radiologic PD assessment	Progressed
3	No progression at the time of data cut-off or withdrawal from study	Date of last adequate radiologic assessment prior to or on date of data cut-off or withdrawal from study	Censored
4	New anticancer treatment started	Date of last adequate radiologic assessment prior to or on date of new anticancer treatment	Censored
5	Progression after more than one missed radiological assessment visit	Date of last adequate radiologic assessment before missed tumor assessments	Censored

* Adequate tumor assessment is a radiologic assessment of CR, PR, SD, non-CR/non-PD or PD as determined by the reviewers.

DOR, TTR, and TTP by BIRC per RECIST v1.1 will be analyzed similarly to PFS. DOR and TTR will be summarized descriptively among all patients with confirmed CR/PR.

7.4.2.4 Health-Related Quality of Life

The EORTC QLQ-C30 (QlQ-C-30) consists of thirty questions that are specific to cancer and cancer treatment (Aaronson NK, et al., 1993; Fayers PM, et al., 2001). It includes a global health status (GHS/Qol) scale that consists of 2 items, and five functional scales measuring Physical (5 items), Role (2 items), Cognitive (2 items), Emotional (4 items), and Social (2 items), three symptom scales measuring Fatigue (3 items), Pain (2 items), and Nausea and Vomiting (2 items) and six single items measuring Dyspnoea, Insomnia, Appetite Loss, Constipation, Diarrhea, and Financial Difficulties. QLQ-C30 scores are based on 4-point Likert scales from 1 = "Not At All" to 4 = "Very Much" with lower scores indicating better HRQoL; except for the GHS/Qol that is scored on a 7-point scale ranging from 1 = "Very Poor" to 7 = "Excellent" with higher scores indicating better health status.

The EORTC QLQ-HCC18 (HCC 18) is the cancer-specific module of QlQ-C30 measuring HCC symptoms. HCC 18 consists of 18 questions and comprises of six symptom scales measuring Fatigue (3 items), Jaundice (2 items), Body Image (2 items), Nutrition (5 items), Pain (2 items), Fever (2 items) and two single items measuring Sex Life and Abdominal Swelling. HCC18 scores are based on the 4-point Likert scale from 1 = "Not at all" to 4 = "Very Much". Lower scores indicate better HRQoL.

The EQ-5D-5L comprises a descriptive module and a Visual Analogue scale (VAS). The descriptive module comprises of 5 dimensions: mobility, self-care, usual activities, pain/ discomfort, and anxiety/depression. Each dimension has 5 levels: 1= no problems, 2 = slight problems, 3 = moderate problems, 4 = severe problems, and 5 = extreme problems. Higher scores indicate lower quality of life. The EQ VAS measures respondent's self-rated health status on a 0 to 100 scale, with 100 = 'the best health you can imagine' and 0 = 'the worst health you can imagine'. Higher scores on VAS indicate higher health status.

EORTC Scoring Derivation

The principle for scoring applies to all scales/scores. The derived scales are obtained from the raw scores as defined in the EORTC manual. If at least half of the items for a scale are answered, then the remaining completed items are used to calculate the score for that scale; however, if more than half are missing, the scale score is set to missing.

Raw Score (RS)

Raw scores are calculated as the average of the items that contribute to the scale:

$$RS=(I_1+I_2+\ldots+I_n)/n$$

Derived Scale (S)

A linear transformation to standardize the raw scores is utilized, so that the scores are ranged from 0 to 100. The derived scales have a more intuitive interpretation: higher scores in functional scales

and the global health status/QoL(GHS/Qol) indicate improvements while higher scores in symptom scales and items indicate deteriorations. The derivation formulae are computed as follows:

Functional scales:

$$S=[1- (RS-1)/range]*100$$

Symptom scales and global health status:

$$S=[(RS-1)/range]*100$$

Refer to Appendix 11.7 and 11.8 for EORTC QLQ HCC18 and EORTC QLQ C30 items and scales. The Index cores are computed as follows using scores of the scales and single items. The higher index score in both HCC18 and QLQ-C30 indicate worse overall HRQoL (Li 2017).

HCC18 index- score=∑ (Scores of Fatigue, Body image, Jaundice, Nutrition, Pain, Fever, Sex Life, Abdominal Swelling) ÷ 8

QLQ-C30 index- score =∑ [(100- Physical functioning), (100- Role functioning), (100- Emotional functioning), (100- Cognitive functioning), (100- Social functioning), (100- GHS/QoL), scores of Fatigue, Nausea/Vomiting, Pain, Dyspnoea, Insomnia, Appetite loss, Constipation, Diarrhea, Financial Difficulty] ÷ 15

All HRQoL measures will be summarized in the ITT analysis set.

A questionnaire module is considered complete if at least one question is answered. The adjusted completion rate is defined as the percentages of patients who completed the questionnaire at each visit divided by the number of patients still in treatment who were expected to complete the questionnaire. Completion rates and adjusted completion rates for the EORTC QLQ C30, EORTC QLQ HCC18, and EQ-5D-5L will be summarized separately at each visit.

The index scores and derived scores (functional scales/symptom scales/single items and the GHS/QoL scale) of QLQ C30 and HCC18 as well as change from baseline at each cycle will be summarized using descriptive statistics. For EQ-5D-5L, descriptive modules will be summarized by visit and dimension in an ordinal scale. The EQ VAS and change from baseline will be summarized by visit in a continuous scale using descriptive statistics. The plots of the mean values and their standard errors at each visit over time by treatment arm will be presented.

In addition, time to deterioration analysis will be performed to compare key quality of life scores between the two treatment groups. Time to deterioration is defined as the time from randomization to the first occurrence of an increase in ≥10 scores (HCC18 index score and global health status/QoL (GHS/QoL) of the QLQ-C30). A deterioration is not counted as an event if a subsequent improvement returned the overall worsening from baseline to less than 10 points. If a patient does not have an event (death or 10% deterioration), they are censored at their last clinic visit at which HRQoL is measured. A nonparametric Kaplan-Meier method will be used to estimate the deterioration curve in each group. The log-rank test and hazard ratio will be provided to show the magnitude of treatment effect and are only used for descriptive purposes.

In addition, a mixed effect model analysis for measuring changes post-baseline will be performed using the GHS/QoL, physical function and fatigue domains of QLQ-C30, and HCC18 index, and

fatigue and pain of HCC 18. Differences in the change from baseline to cycle 4 and cycle 6 between arms of the aforementioned parameters will be assessed in the mixed models which include baseline score, stratification factors, treatment arm, visit, and treatment arm by visit interaction as fixed effects and visit as a repeated measure with an unstructured covariance matrix. The compound symmetry covariance matrix will be used if there is a convergence issue for unstructured covariance matrix.

7.4.2.5 ORR, PFS, DOR, and TTP assessed by the Investigator

ORR, PFS, DOR, TTR ,and TTP assessed by Investigator will be analyzed similarly to the approach used in the assessment by BIRC.

7.4.2.6 Disease Control Rate (DCR) and Clinical Benefit Rate (CBR) by BIRC and Investigators

Disease control rate (DCR) is defined as the proportion of patients whose best overall response (BOR) is CR, PR, or SD. Clinical benefit rate (CBR) is defined as the proportion of patients who have CR, PR, or SD of $\geq$ 24 weeks in duration.

Both DCR and CBR will be analyzed similarly to ORR in the ITT analysis set.

Response assessment will be compared between the BIRC and investigators. Concordance in response assessments will be reported as the numbers and percentages of responders (CR or PR) and non-responders, as assessed by both the BIRC and the investigators. Discordance in response assessments will be reported as the numbers and percentages of patients whose best overall response assessments are different between the BIRC and the investigators. Cases in which one of these assessments is missing will be considered as discordance.

Waterfall plots will be provided for the maximum tumor shrinkage based on target lesion. In addition, patients will be marked out in the plot for those who had tumor reduction in target lesion assessments but contradicted the PD results in overall response due to new-lesion or non-target lesion. The maximum tumor shrinkage based on target lesion used in the plots will be listed. These analyses will be performed based on RECIST1.1.

7.4.3 Subgroup Analyses

To determine if the treatment effect is consistent across various subgroups, the hazard ratio estimates of OS and its 95% CI will be estimated and plotted within each category of the following variables:

- geography (Asia excluding Japan vs Japan vs Rest of World)
- macrovascular invasion and/or extrahepatic spread (present or absent)
- macrovascular invasion (present or absent)
- extrahepatic spread (present or absent)
- age (< 65 vs $\geq$65 years)
- gender (Female vs Male)
- race (Asian vs White vs Other)

- ECOG PS (0 vs 1)
- hepatitis virus infection (HBV vs HCV vs Non-viral)
- BCLC stage (Stage B vs Stage C)
- loco-regional procedure (yes or no)
- AFP subgroups: < 200 ug/L, >= 200 ug/L, < 400 ug/L, >=400 ug/L

Odds ratio of ORR and hazard ratio of PFS will be summarized in the subgroups as well.

Country-specific subgroups may also be summarized per local regulatory requirements.

7.4.4 Post-Study Anti-Cancer Therapy

Post-study anti-cancer therapy is defined as the anti-cancer therapy started after the last dose date of study drug. A summary of number and pecentage of patients who received subsequent systematic anticancer therapy/immune checkpoint inhibitors (single treatment), and combination therapy of immune checkpoint inhibitors and tyrosine kinase inhibitors will be provided by arm based on the ITT analysis set.

Patient data listings of post-study anti-cancer therapy, procedure, radiotherapy, embolization, or surgery will be provided.

7.5 Safety Analyses

Safety will be assessed by monitoring and recording of all AEs graded by NCI-CTCAE v4.03. Laboratory values (e.g., hematology, clinical chemistry), vital signs, ECGs, and physical exams, will also be used in determining safety. Descriptive statistics (e.g., n, mean, standard deviation, median, Q1, Q3, minimum, maximum for continuous variables; n [%] for categorical variables) will be used to analyze all safety data in the safety analysis set.

7.5.1 Extent of Exposure

The following exposure parameters will be summarized with descriptive statistics for each study drug. One cycle is defined as 21 days of treatment. Specifically:

Treatment duration (TD) for tislelizumab: The treatment duration will be calculated as (last date of exposure – date of first dose + 1)

- If patients discontinued treatment (with non-missing EOT date), using min (cutoff date, death date, last dose date + 20) as the “last date of exposure”
- otherwise if patient has treatment ongoing, using cutoff date as the “last date of exposure” for calculation of TD

Treatment duration for sorafenib: date of the last dose of sorafenib – date of first dose of sorafenib + 1). If the first dose starts in PM or the last dose stops in AM, only half day will be counted.

Total Cumulative Dose for tislelizumab and sorafenib = sum (all actual dosages per administration at all visits prior to the cut-off date).

Actual Dose Intensity (ADI) for tislelizumab (mg/cycle) = 21*total cumulative dose (mg) / (last dose date prior to cut off date + 21 - first dose date).

Actual Dose Intensity for sorafenib (mg/cycle) = (total cumulative dose (mg) / TD for sorafenib) * 21.

Planned Dose Intensity for tislelizumab (mg/cycle) = 200 mg/cycle.

Planned Dose Intensity for sorafenib (mg/cycle) = 16800 mg/cycle.

Relative Dose Intensity (%) = Actual Dose Intensity/Planned dose intensity *100%

Number of cycles received is defined as the sum of number of cycles with at least one non-missing doses (dose>0). If a patient takes at least one dose of sorafenib in the 21-day period, the cycle will be counted. For tislelizumab, if a patient did not receive tislelizumab IV during the 21-day period, this cycle will not be counted.

Average cycle length of tislelizumab = (date of the last dose of tislelizumab – date of first dose + 21) / the number of cycles received.

Average cycle length of sorafenib = treatment duration for sorafenib / the number of cycles received.

The number of patients with dose reductions (sorafenib only), dose omissions, dose delays, dose interruptions, and treatment discontinuation and their reasons will be summarized by counts and percentages according to study drug. In addition, frequency of dose reductions (sorafenib only), dose omission, dose delays and dose interruptions will be summarized by categories (0, 1, ≥2).

Patient data listings will be provided for all dosing records, and for the above calculated summary statistics.

7.5.2 Adverse Events

The AE verbatim descriptions (Investigator's description from the eCRF) will be classified into standardized medical terminology using Medical Dictionary for Regulatory Activities (MedDRA). Adverse events will be coded to the latest MedDRA version lower-level term closest to the verbatim term. The linked MedDRA System Organ Class (SOC) and Preferred Term are also classified. All adverse event summaries are based on safety analysis set.

In this trial, a TEAE is defined as an AE that had an onset date or a worsening in severity from baseline (pretreatment) on or after the first dose of study drug up to 30 days following study drug discontinuation or initiation of new anticancer therapy, whichever occurs first. Only those AEs that were treatment-emergent will be included in summary tables. All AEs, treatment-emergent or otherwise, will be presented in patient data listings. COVID-19 related adverse events will be summarized separately.

It is noteworthy that the definition of TEAE in the protocol is different from the one in this SAP, while the definition of imAE remains the same. The update of the TEAE window streamlines the TEAE derivation so all TEAEs can be identified programmatically instead of relying on manual medical review. Any imAE occurring outside of the above mentioned TEAE window will not be classified as TEAEs. All imAEs will be reported separately.

The incidence of TEAEs will be reported as the number (percentage) of patients with TEAEs by System Organ Class and Preferred Term. A patient will be counted only once by the highest severity grade within a System Organ Class and Preferred Term, even if the patient experienced more than one TEAE within a specific System Organ Class and Preferred Term. The number (percentage) of patients with TEAEs will also be summarized by relationship to the study drug.

7.5.2.1 Treatment Emergent Adverse Event

An overall summary of TEAEs will summarize the number (%) of patients with

- At least one TEAE
- At least one TEAE with NCI-CTCAE grade ≥3
- At least one treatment-related TEAE
- At least one serious TEAE
- At least one TEAE leading to death
- At least one TEAE leading to discontinuation of study drug
- At least one TEAE leading to dose modification of study drug
- At least one infusion-related reaction
- At least one infusion-related reaction with NCI-CTCAE grade ≥3

Summaries of the following TEAEs will be provided:

- All TEAEs
    - All TEAEs by SOC
    - All TEAEs by SOC and PT
    - Most frequently reported (incidence ≥10% in any treatment arm) TEAE by SOC and PT
    - Treatment-related TEAE by SOC and PT
    - Most frequently reported (incidence ≥10% in any treatment arm) Treatment-related TEAE by SOC and PT
- Serious TEAEs by SOC and PT
    - Most frequently reported (incidence ≥ 5% in any treatment arm) serious TEAE by SOC and PT
    - Treatment-related Serious TEAE by SOC and PT
- TEAEs with NCI-CTCAE grade ≥3 by SOC and PT
    - Treatment-related TEAE with NCI-CTCAE grade ≥ 3 by SOC and PT
- Most frequently reported (incidence ≥ 5 % in any treatment arm) TEAE with NCI-CTCAE grade ≥ 3 by SOC and PT.

- TEAEs leading to death by SOC and PT
  - Treatment-related TEAE Leading to Death by SOC and PT
- TEAEs leading to treatment discontinuation by SOC and PT
  - Treatment-related TEAE Leading to Treatment Discontinuation by SOC and PT
- TEAEs leading to dose modification by SOC and PT
  - Treatment-related TEAE Leading to Dose Modification by SOC and PT

7.5.2.2 Immune-mediated Adverse Event

Immune-mediated adverse events are of special interest and summarized by category within a pre-defined list in Appendix 11.6. The identification of immune-mediated adverse events is described in the immune-mediated adverse event charter. All immune-mediated adverse events occurring up to 90 days after the last dose of tislelizumab will be reported.

Summaries of the following incidences of immune-mediated adverse events will be provided:

- Immune-mediated adverse events by category and maximum severity
- Immune-mediated adverse events with NCI-CTCAE grade ≥3 by category
- Immune-mediated adverse events leading to treatment discontinuation by category
- Immune-mediated adverse events leading to death by category
- Immune-mediated adverse events leading to dose modification by category
- Immune-mediated adverse events treated with systematic corticosteroid by category

7.5.2.3 Infusion-related Adverse Event

The PT list of infusion-related reactions (IRRs) included fever/pyrexia, chills/rigor, nausea, pruritus, angioedema, hypotension, headache, bronchospasm, urticaria, rash, vomiting, myalgia, dizziness, hypertension, acute respiratory distress syndrome, myocardial infarction, ventricular fibrillation, cardiogenic shock, IRR, and transfusion reaction. The following process was used for the final determination of IRRs:

a. The investigator/site must have checked the IRR box on the AE pages in the CRF.

b. The event term must have matched (or been equivalent) to the IRR terms listed above (such as fever or chills), events that happened concurrently with one of the terms on this list (such as fever + back pain + chest pain – all would be included).

c. Only events that started on the day of an infusion or the day after an infusion were included.

For IRRs, a summary of incidence by SOC, PT and maximum severity will be provided, sorted by descending order of incidence within each SOC and PT based on tislelizumab column. Summaries of IRRs, IRRs with NCI-CTCAE grade ≥3 and IRRs leading to treatment discontinuation will also be provided by PT only, in descending order.

7.5.3 Laboratory Values

Clinical laboratory (e.g., hematology, serum chemistry, coagulation, thyroid function, and urinalysis) values will be evaluated for each laboratory parameter by patient. Abnormal laboratory values will be flagged and identified as those outside (above or below) the normal range. Reference (normal) ranges for laboratory parameters will be included in the clinical study report for this protocol. Descriptive summary statistics (e.g., n, mean, standard deviation, median, minimum, maximum for continuous variables; n [%] for categorical variables) for laboratory parameters and their changes from baseline will be calculated. Laboratory values will be summarized by visit and by worst postbaseline visit.

Laboratory parameters that are graded in NCI-CTCAEv.4.03 will be summarized by NCI-CTCAE grade. In the summary of laboratory parameters by NCI-CTCAE grade, parameters with NCI-CTCAE grading in both high and low directions (e.g., calcium, glucose, magnesium, potassium, sodium) will be summarized separately. Shift tables will be used to summarize the grade change from baseline to worst post baseline value with counts and percentages.

Hy's Law for liver injury will also be summarized.

Table 3 Clinical Laboratory Assessments

Serum Chemistry	Hematology	Coagulation	Urinalysis	Thyroid Function
Alkaline phosphatase Alanine aminotransferase Aspartate transaminase Albumin Direct bilirubin Total bilirubin Blood urea nitrogen or urea Creatinine Lactate dehydrogenase Total protein Potassium Sodium Calcium	Hemoglobin Hematocrit Platelet count WBC count Neutrophil count Lymphocyte count	Prothrombin time Activated Partial Thromboplastin Time International Normalized Ratio	Glucose Protein Ketones Blood 24-hour protein Random urine protein to creatinine ratio	TSH Free T3 Free T4

Magnesium Phosphate Creatine kinase Creatine kinase cardiac muscle isoenzyme (CK-MB)				

Abbreviations: WBC = white blood cell. TSH = Thyroid stimulating hormone

7.5.4 Vital Signs

Descriptive statistics for vital sign parameters (systolic and diastolic BP, heart rate, temperature, weight) and changes from baseline will be presented by visit. For tislelizumab, the change from post-dose (end of infusion) to pre-dose also needs to be summarized for all vital sign parameters except for height and weight. Vital signs will be listed by patient and visit.

7.5.5 Electrocardiograms (ECG)

12-lead ECG recordings are required at Screening, Safety Follow-up, and as clinically indicated. Patient listing of ECG will be provided for all ECG recordings.

The actual value and the change from baseline for QTcF intervals will be summarized by visit and treatment group using descriptive statistics.

Abnormal post-baseline QTcF results will be summarized with the following categories: increase of >30 msec, increase of > 60 msec, value of > 450 msec, value of > 480 msec, value of > 500 msec for each visit by treatment group.

Shift table of ECG results (normal, abnormal not clinically significant, abnormal clinically significant) will be provided for change from baseline to worst post baseline visit by treatment group.

7.5.6 ECOG

ECOG performance status will be summarized by treatment arm and by visit.

A shift table from baseline to worst post-baseline in ECOG performance score will be summarized.

Patient listing of ECOG will be provided for all ECOG findings.

7.5.7 Antidrug Antibody

Samples to assess anti-tislelizumab antibodies will be collected only in patients randomized to receive tislelizumab and in sites that are able to adequately perform sampling, handling, and processing procedures outlined in the laboratory manual.

ADA attributes:

- **Treatment-boosted ADA** is defined as ADA positive at baseline that was boosted to a 4-fold or higher-level following drug administration.
- **Treatment-induced** ADA is defined as ADA negative at baseline and ADA positive post-baseline.
- **Transient ADA response** is defined as treatment-emergent ADA detected only at 1 time point during treatment or follow-up, excluding last time point; or detected at 2 or more time points during treatment or follow-up, where the first and last positive samples are separated by less than 16 weeks and the last time point is negative.
- **Persistent ADA response** is defined as treatment-induced ADA detected at 2 or more time points during treatment or follow-up, where the first and last ADA positive samples are separated by 16 weeks or longer; or detected only in the last time point or at a time point less than 16 weeks before a negative last sample.
- **Neutralizing ADA** is defined as ADA that inhibits or reduces the pharmacological activity.

ADA response endpoints:

- **ADA incidence** is defined as sum of treatment-induced and treatment-boosted ADA-positive patients as a proportion of the ADA evaluable population.
- **ADA prevalence** is defined as proportion of all patients that are ADA positive, including pre-existing ADA, at any time point.

The immunogenicity results will be summarized using descriptive statistics by the number and percentage of patients who develop detectable ADA. The incidence of positive ADA and neutralizing ADA will be reported for evaluable patients. The effect of immunogenicity on PK, efficacy and safety may be evaluated if data allow.

7.5.8 Other Safety Measurements

Other safety measurements including pulmonary function, ophthalmology exam, digital pulse oximetry, and abnormal physical exam findings, will be listed by patient and visit.

7.6 Pharmacokinetic Analyses

Pharmacokinetic samples will be collected in this study as outlined in appendix 1 of the protocol, and only from patients randomized to receive BGB-317 in sites that are able to adequately perform PK sampling, handling, and processing procedures as outlined in the Laboratory Manual.

Tislelizumab post-dose and trough serum concentration data (Ctrough) will be tabulated and summarized by visit/cycle at which these concentrations are collected. Descriptive statistics will include geometric means, geometric CV%, medians, ranges, and standard deviations, as appropriate.

Additional PK analyses, including population PK analyses and exposure-response (efficacy, safety endpoints) analyses may be conducted as appropriate and the results of such analysis may be reported separately from the clinical study report. Concentrations of tislelizumab will be summarized descriptively.

7.7 OTHER ANALYSIS

Distribution of PD-L1 expression will be examined in the ITT analysis set. Potential association between PD-L1 expression and tislelizumab treatment effect over sorafenib will be explored.

Other potential predictive markers will be assessed.

8 INTERIM ANALYSIS

There will be one interim analysis of OS superiority. The interim analysis will be performed when approximately 403 deaths (80% of the expected number of approximately 504 death events) between the 2 treatment arms are observed. It is estimated that it will take approximately 33.9 months to observe 403 events. The final analysis of OS will take place after 504 deaths are observed. The upper (efficacy) boundary is based on the O'Brien-Fleming boundary approximated by the Hwang-Shih-DeCani spending function by setting gamma parameter at -4. Stopping boundaries (p-value and Z score) of the superiority test for OS at the interim and final analyses are shown in Table 1. These boundaries will be updated from the actual death events observed at the time of the interim analysis and final planned number of deaths, using the alpha-spending function.

Table 1 Stopping Boundaries and Approximate Hazard Ratio Threshold for Interim and Final Analyses of Superiority Test for Overall Survival

	Time (months)	#Deaths	p-value (Z score) for Efficacy	Approximate Hazard Ratio for Efficacy
Interim analysis	33.9	403	< 0.0110 (> 2.29)	< 0.7958
Final analysis	46.6	504	< 0.022 (> 2.01)	< 0.8358

9 CHANGES IN THE PLANNED ANALYSIS

Not applicable.

10 REFERENCES

Aaronson NK, Ahmedzai S, Bergman B, *et al.* The European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute*. **85:** 365-76, 1993.

Common Terminology Criteria for Adverse Events (CTCAE). Version 4.03. United States Department of Health and Human Services, National Institutes of Health, National Cancer Institute, Washington, DC, USA, June 14, 2010.

Eisenhauer EA, Therasse P, Bogaerts J Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eur J Cancer. 2009;45:228-47.

Fayers PM, Machin D. *Quality of Life: Assessment, Analysis and Interpretation*. J Wiley & Sons Ltd, Chichester, 2000.

Guidance for Industry: Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. Washington, DC: United States Food and Drug Administration; 2007.

Kaplan E, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc. 1958;53:457-81.

Li L, Mo FKF, Chan SL, et al. Prognostic values of EORTC QLQ-C30 and QLQ-HCC18 index-scores in patients with hepatocellular carcinoma – clinical application of health-related quality-of-life data. BMC Cancer. 2017; 17:8.

Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. Stat Med. 1998;17:2815-34.

Robins JM, Tsiatis A. Correcting for non-compliance in randomized trials using rank-preserving structural failure time models. Communications in Statistics. 1991;20: 2609-31.

Rothmann M, Li N, Chen G, Chi GYH, Temple RT, Tsou HH. Design and analysis of non-inferiority mortality trials in oncology. Stat Med. 2003;22:239-64.

Uno H, Claggett B, Tian L, Inoue E, Gallo P, Miyata, et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis J Clin Oncology. 2014; 32(22):2380-5

Yau et al. ESMO 2019, Abstract # 6572

11 APPENDIX

11.1 IMPUTE PARTIAL DATES FOR CONCOMITANT MEDICATION

When the start date or end date of a medication/therapy/procedure is partially missing, the date will be imputed to determine whether the medication/therapy/procedure is prior or concomitant. The following rules will be applied to impute partial dates for medications.

If start date of a medication/therapy/procedure is partially missing, impute as follows:

- If both month and day are missing, then set to January 01
- If only day is missing, then set to the first of the month
- If the imputed start date > death date, then set to death date

If end date of a medication/therapy/procedure is partially missing, impute as follows:

- If both month and day are missing, then set to December 31
- If only day is missing, then set to last day of the month
- If the imputed end date > death date, then set to death date

If the year of start date or year of end date of a medication/therapy/procedure is missing, or the start date or end date is completely missing, do not impute.

11.2 IMPUTE PARTIAL DATES FOR ADVERSE EVENTS

If year of the start date is missing or start date is completely missing, do not impute. Impute AE end date first if both AE start date and end date are partially missing.

If end date of an adverse event is partially missing, impute as follows:

- If both month and day are missing, then set to December 31
- If only day is missing, then set to last day of the month
- If the imputed end date > death date, then set to death date

If year of the end date is missing or end date is completely missing, do not impute. If start date of an adverse event is partially missing, impute as follows:

- If both month and day are missing and year = year of treatment start date, then set to treatment start date
- If both month and day are missing and year ≠ year of treatment start date, then set to January 01
- If day is missing and month and year = month and year of treatment start date, the set to treatment start date
- If day is missing and month and year ≠ month and year of treatment start date, the set to first of the month
- If the imputed AE start date is after AE end date (maybe imputed), then update AE start date with AE end date as final imputed AE start date. If the imputed end date >

death date, then set to death date.

11.3 IMPUTE PARTIAL DATES FOR SUBSEQUENT ANTI-CANCER SURGERY/PROCEDURE

When the start date of subsequent anti-cancer therapy is partially missing, the following rules will be applied to impute partial dates.

If start date of is partially missing, impute as follows:

- If both month and day are missing, then set to December 31
- If only day is missing, then set to last day of the month
- If the imputed start date > death date, then set to death date

If year of the start date is missing, do not impute. If imputed start date is after study discontinuation date, then set to study discontinuation date.

11.4 IMPUTE PARTIAL DATES FOR PRIOR ANTI-CANCER THERAPY (DRUG, SURGERY/PROCEDURE, RADIOTHERAPY)

The following rules will be applied to impute partial dates such as initial diagnosis date, initial BCLC staging date, relapse date, therapy date (start/end date), or surgery date etc.

- If start date of a disease history or prior therapy is partially missing, impute as follows:
- If both month and day are missing, then set to January 01
- If only day is missing, then set to the first of the month

If the imputed start date > first dose date then set to first dose date – 1

If end date of a disease history or prior therapy is partially missing, impute as follows:

- If both month and day are missing, then set to December 31
- If only day is missing, then set to last day of the month
- If the imputed end date > first dose date, then set to first dose date -1

If the year of start date or year of end date of a medication/therapy/procedure is missing, or the start date or end date is completely missing, do not impute. If imputed start date/end date is after randomization date - 14 , then set to randomization date - 14.

11.5 Progression Free Survival Censoring Rule

No.	Situation	Date of Progression or Censoring	Primary Analysis	Sensitivity Analysis 1	Sensitivity Analysis 2
1	No baseline tumor assessments and without death within 10 weeks after randomization	Randomization date	Censored	Censored	Censored
2	Progression documented between scheduled visits	Date of first radiologic PD assessment	Progressed	Progressed	Progressed***
3	No progression at the time of data cut-off or withdrawal from study	Date of last adequate radiologic assessment prior to or on date of data cut-off or withdrawal from study	Censored	Censored	Censored
4	New anticancer treatment started	Date of last adequate radiologic assessment prior to or on date of new anticancer treatment	Censored	-	Censored
5	Death before first PD assessment	Date of death	Progressed	Progressed	Progressed***
6	Death between adequate assessment visits*	Date of death	Progressed	Progressed	Progressed***
7	Death or progression after more than one missed visit**	Date of last adequate radiologic assessment before missed tumor assessments	Censored	Censored	Progressed or died
8	Discontinued due to Clinical PD but no documented PD	Date of clinical progression	-	-	-

* Adequate tumor assessment is a radiologic assessment of CR, PR, SD, non-CR/non-PD or PD as determined by the reviewers.

** More than one missed visit is identified in Section Identifying two missing tumor assessment below

*** Progression date for PFS event will be the earliest date of events defined in 2,4,5,6

11.6 Immune-mediated Adverse Event Category List

Category
Immune-mediated adrenal insufficiency
Immune-mediated anaemia
Immune-mediated colitis
Immune-mediated hepatitis
Immune-mediated hyperthyroidism
Immune-mediated hypothyroidism
Immune-mediated myocarditis
Immune-mediated myositis/rhabdomyolysis
Immune-mediated nephritis and renal dysfunction
Immune-mediated nervous system disorder
Immune-mediated ocular disorder
Immune-mediated pancreatitis
Immune-mediated pituitary dysfunction
Immune-mediated pneumonitis
Immune-mediated skin adverse reaction
Immune-mediated thrombocytopenia
Immune-mediated thyroiditis
Immune-mediated type 1 diabetes mellitus
Other immune-mediated reactions

11.7 HCC18 Specific Symptoms Scales

	Scale	Number of items	Item range	HCC18 Item Numbers
Symptom Scales				
Fatigue	Fati	3	3	45-47
Jaundice	Jaun	2	3	36-37
Body Image	BI	2	3	33, 35
Nutrition	Nutn	5	3	31, 32, 42-44
Pain	Pain	2	3	38-39
Fever	Fev	2	3	40-41
Single Items				
Sex Life	Sx	1	3	48
Abdominal Swelling	Ab	1	3	34

11.8 Scoring of QLQ-C30

	Scale	Number of items	Item range	Item Numbers
Global health status/ QoL	QL2	2	6	29,30
Functional Scales				
Physical functioning	PF2	5	3	1, 2, 3, 4, 5
Role functioning	RF2	2	3	6, 7
Emotional functioning	EF	4	3	21, 22, 23, 24
Cognitive functioning	CF	2	3	20, 25
Social functioning	SF	2	3	26, 27
Symptom Scales/ items				
Fatigue	FA	3	3	10, 12, 18
Nausea and vomiting	NV	2	3	14, 15
Pain	PA	2	3	9, 19
Dyspnoea	DY	1	3	8
Insomnia	SL	1	3	11
Appetite loss	AP	1	3	13
Constipation	CO	1	3	16
Diarrhoea	DI	1	3	17
Financial Difficulties	FI	1	3	28

11.9 List of Chinese herbal medicine or chinese patent medicines which have effect of control cancer or boost immunity

The following table lists those medications that require a 14-day wash-out and should be prohibited during the study:

Drug Name (Chinese)	Drug Name (English)
Rg3 参一胶囊	Ginsenoside-Rg3 capsule
养正消积胶囊	Yangzheng Xiaoji Jiaonang
化癥回生口服液	Huazheng Huisheng Koufuye
十全大补汤	Juzentaihoto
华蟾素注射液	Cinobufacini/Huachansu injection
华蟾素片/胶囊	Cinobufacini/Huachansu Pian/Capsules
博尔宁胶囊	Boerning capsule
去甲斑蝥素片	Norcantharidin Pian
参丹散结胶囊	Shendan Sanjie Jiaonang
参芪扶正注射液	Shengqi Fuzheng Zhusheye
参莲胶囊/颗粒	Shen Lian Jiao Nang/Ke Li
吗特灵注射液	Ma Te Ling injection
回生口服液	Hui Sheng Kou Fu Ye
复方斑蝥胶囊	Fufang Banmao Jiaonang
复方红豆杉胶囊	Fufang Hongdoushan Jiaonang
复方苦参注射液	Fufang Kushen Zhusheye
天仙胶囊	Tian Xian capsule
奇宁注射液	Qining injection
威麦宁胶囊	Weimaining Jiao Nang
安尔欣注射液	Anerxin/Ginseng polysaccharide injection
安康欣胶囊	Ankangxin Jiaonang
安替可胶囊	Antike capsule
岩舒注射液	Yanshu injection
平消片/胶囊	Ping Xiao Pian/Jiao Nang
康力欣胶囊	Kanglixin Jiaonang

Drug Name (Chinese)	Drug Name (English)
康艾注射液	Kang'ai Zhusheye
康莱特注射液	Kanglaite Injection
康莱特软胶囊	Kanglaite Soft Capsules
慈丹胶囊	CIDAN Capsule
槐耳颗粒	Huaer Keli
海生素注射液	Haishengsu injection
消癌平丸/片/胶囊/颗粒	Xiaoaiping Wan/Pian/Jiao Nang/Ke Li
消癌平注射液	Xiaoaiping Zhusheye
牛黄醒消丸	Niuhuang Xingxiao pill
猪苓多糖注射液	Polyporus polysaccharide injection
白花蛇舌草注射液	Hedyotis Dissusa wild injection
紫龙金片	Zi Long jin pian
肝复乐片/胶囊	Ganfule Jiaonang / GFL tablet
肿节风片	Zhongjiefeng tablet
胃复春片	Weifuchun pill
艾迪注射液	Ai Di Zhu She Ye
芪珍胶囊	Qizhen Jiaonang
莪术油注射液	Zedoary turmeric oil injection
金复康口服液	Kanglixin Jiaonang
金蒲胶囊	Jinpu capsule
金龙胶囊	Jinlong Capsules
香菇多糖	Lentinan
鸦胆子油乳注射液	Yadanzi/Brucea javanica Youru Zhusheye
鸦胆子油软胶囊/口服乳液	Yadanziyou Ruan jiao nang/Kou Fu Ru Ye

Terminology list: Pian = tablet, Jiao Nang/Jiaonang = capsule, Ke Li/Keli = granules, Zhue She ye/Zhusheye = injections, Kou Fu Ye/koufuye = oral liguid, Wan = Pill/bolus, He Ji/Heji = mixture, Gao = ointment
NOTE: These list of Chinese herbal medicines or Chinese patent medicines are provided as examples and are not intended to be all-inclusive.

11.10 Safety Run-In Sub-study in Japanese Patients

11.10.1 Study Design

This is an open-label, multicenter, non-randomized Phase 1 clinical trial in Japanese patients with unresectable HCC who have previously received standard of care treatment. Screening can last up to 28 days and treatment can continue until the Investigator considers the patient is no longer benefiting from tislelizumab, toxicity, or voluntary withdrawal of study treatment. Approximately 10 patients were enrolled to address the PK profile of tislelizumab in Japanese patients. A Safety Follow-up phase will occur up to 30 days following last study treatment or initiation of new cancer therapy, whichever occurs first, for any AEs, and up to 90 days following last dose of tislelizumab for irAEs, regardless of whether or not the patient starts a new anticancer therapy. Survival follow-up information will be collected approximately every 3 months (± 14 days) after the Safety Follow-up Visit until death, loss to follow-up, withdrawal of consent, or study termination by Sponsor. The first 6 patients will remain hospitalized during the first week of the first cycle of treatment.

11.10.2 Study Assessment

Study assessments and procedures will be similar to those in the main study (Sections 2.2), with the exception that patients will not be randomized for the substudy and there is no second treatment arm. In addition, PK assessments have been increased in this substudy for thorough characterization of PK after single dose and at steady state in Japanese patients. Patients will be closely monitored for safety and tolerability throughout the study.

Patients will be evaluated for any AEs and serious adverse events (SAEs) occurring up to 30 days after the last dose of study drug (all severity grades, per NCI-CTCAE v4.03) or initiation of new anticancer therapy, whichever occurs first, and immune-related AEs (irAEs) occurring up to 90 days after the last dose of study drug, regardless of whether or not the patient starts a new anticancer therapy. All study drug-related SAEs will be recorded by the Investigator after treatment discontinuation until patient death or loss to follow-up, whichever occurs first.

11.10.3 Analysis Set

Safety Analysis Set: Includes all patients who received at least one dose of study drug and is the primary analysis set used for all safety and efficacy analyses.

PK Analysis Set: Includes all patients who receive at least one dose of tislelizumab per the protocol, for whom any post-dose PK data are available.

11.10.4 Primary Analyses

11.10.4.1 Safety Analyses

As described in the main study (Section 7.5), safety will be assessed by monitoring and recording of all AEs graded by NCI-CTCAE v4.03. Laboratory values (e.g., hematology, clinical chemistry coagulation, thyroid function, and urinalysis), vital signs, ECGs, and physical examinations will also be used in determining safety. The incidence of treatment-emergent adverse events (TEAEs) will be reported as the number (percentage) of patients with TEAEs by System Organ Class and Preferred Term. Descriptive summary statistics (e.g., n, mean, standard deviation, median, minimum, maximum for continuous variables; n [%] for categorical

variables) and changes from baseline will be determined for laboratory parameters and vital signs.

11.10.4.2 Pharmacokinetic Analyses

The PK analysis will use noncompartmental methods to calculate the following PK parameters, as appropriate and allowed by data: including but not limited to area under the concentration-time curve from Day 0 to Day 21 (area under the concentration-time curve from Day 0 to Day 21 [AUC_{0-21d}], C_{max} , time to maximum plasma concentration [T_{max}], trough serum concentration [$C_{through}$], elimination half-life ($t_{1/2}$), clearance (CL) and volume of distribution (Vd). Mean serum tislelizumab concentration data and PK parameters will be tabulated and summarized by visit/cycle at which these data are available. Concentrations of tislelizumab will be summarized descriptively. Descriptive statistics will include geometric means, medians, ranges, and standard deviations, as appropriate.

11.10.4.3 Other Analyses

The demographics, baseline characteristics and disease history, prior systematic anti-cancer therapy, prior radiotherapy, prior surgeries/procedures, prior and concomitant medication will also be summarized.

11.10.5 Secondary Analyses

Efficacy evaluations per RECIST v1.1 (i.e., ORR, PFS and DOR) will be summarized to explore the preliminary anticancer activities in Japanese patients

The definition of efficacy endpoints PFS, ORR, and DOR are same as section 5.1.1 and 5.1.2. Efficacy measurements ORR, PFS, and DOR, will be listed for all patients received at least one dose in Japanese sub-study.

Waterfall plot of maximum tumour shrinkage per patient will be presented.

Patient data of immunogenic responses to tislelizumab will be listed.

11.10.6 Exploratory Analyses

Assessments of the correlations between drug exposure and response (efficacy and safety endpoints) will be made. Results of such exploratory analysis may be reported separately from the clinical study report (CSR).