STATISTICAL ANALYSIS PLAN

Study Protocol Number:	BGB-A317-208
Study Protocol Title:	A Phase 2, Open-label, Multicenter Study to Investigate the Efficacy, Safety, and Pharmacokinetics of the Anti-PD-1 Monoclonal Antibody BGB-A317 in Patients with Previously Treated Hepatocellular Unresectable Carcinoma
Date:	April 8, 2020
Version:	1.2

SIGNATURE PAGE

Author: [redacted]

Signature

Date: _____

Approval

[redacted]

Signature

Date: _____

[redacted]

Signature

Date: _____

[redacted]

Signature

Date: _____

[redacted]

Signature

Date: _____

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Term
ADA	Anti-drug Antibody
ADI	Actual Dose Intensity
AE	Adverse Event
BCLC	Barcelona Clinic Liver Cancer
BGB-A317	Code name for Monoclonal Antibody BGB-A317 (tislelizumab)
BID	Twice daily
BMI	Body Mass Index
BOR	Best Overall Response
CBR	Clinical Benefit Rate
CI	Confidence Interval
CR	Complete Response
CRF	Clinical Report Form
CTCAE	Common Terminology Criteria for Adverse Events
C_{trough}	Minimum Observed Plasma Concentration
DCR	Disease Control Rate
DOR	Duration of response
EBRT	External Beam Radiation Therapy
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic Case Report Form
EFF	Efficacy Evaluable Set
EQ-5D	European Quality of Life 5-Dimensions
EORTC QLQ-HCC 18	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Hepatocellular Carcinoma 18 Questions
EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30
FDA	Food and Drug Administration
HBV	Hepatitis B Virus
HCC	Hepatocellular Carcinoma
HCV	Hepatitis C Virus
HRQoL	Health Related Quality of Life
irAE	Immune-Related Adverse Event
IRC	Independent Review Committee
IV	Intravenous(ly)
MedDRA	Medical Dictionary for Regulatory Activities
NCI-CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
ORR	Objective Response Rate
OS	Overall Survival

PD	Progressive Disease
PD-1	Programmed Cell Death-1
PD-L1	Program Death Ligand-1, Programed Death Receptor Ligand-1, Programed Death-1 Ligand-1
PFS	Progression Free Survival
PK	Pharmacokinetics
PR	Partial Response
PT	Prothrombin Time or Preferred Term
Q3W	Once Every 3 Weeks
QTc	QT Interval Corrected for Heart Rate
RDI	Relative Dose Intensity
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	Serious Adverse Event
SAF	Safety Analysis Set
SAP	Statistical Analysis Plan
SD	Stable Disease, Standard Deviation
SI	System International
SOC	System Organ Class
TACE	Transarterial Chemoembolization
TEAE	Treatment-Emergent Adverse Event
TIL	Tumor-infiltrating lymphocytes
WBC	White Blood Cell
WHO-DD	World Health Organization Drug Dictionary

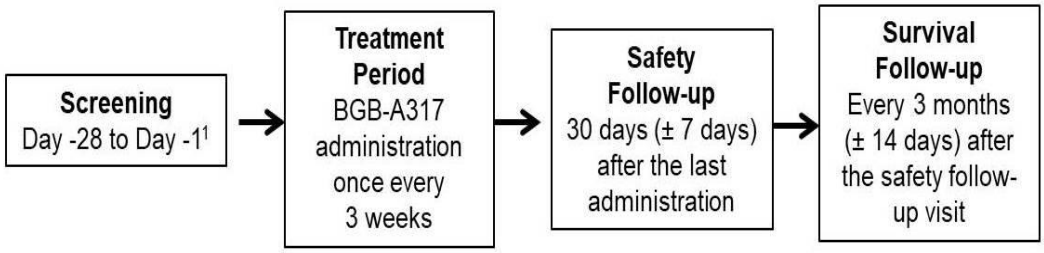
1 INTRODUCTION

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 Protocol A317-208 entitled: "A Phase 2, Open-label, Multicenter Study to Investigate the Efficacy, Safety, and Pharmacokinetics of the Anti-PD-1 Monoclonal Antibody BGB-A317 in Patients with Previously Treated Hepatocellular Unresectable Carcinoma". The focus of this SAP is for the planned analyses specified in the study protocol amendment version 2.0 dated 25 June 2018.

2 STUDY OVERVIEW

Study Design:

This is a Phase 2, multicenter, open-label study in adults with previously treated unresectable HCC. Approximately 228 patients will receive tislelizumab (BGB-A317) 200 mg intravenously (IV) every 3 weeks (Q3W). The study design schema is as follows:



[1] Screening assessments will be completed within 28 days prior to the first dose of the study drug. If laboratory tests at screening are not performed within 7 days prior to the administration of study drug on Cycle 1 Day 1, chemistry, hematology, and coagulation tests should be repeated and reviewed before study drug administration.

After completing all Screening activities, patients confirmed by the sponsor to be eligible will receive open-label treatment with tislelizumab 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, whichever should occur first. Treatment beyond initial disease progression (as assessed by the investigator per RECIST v1.1) is permitted, provided that the patient meets protocol-specified criteria. Of the 228 patients, at least 100 patients will be enrolled who have had no more than 1 line of prior systemic therapy and at least 100 patients will be enrolled who have had at least 2 lines of prior systemic therapy.

Study Assessments:

Radiological assessment of tumor-response status should be performed every 6 weeks in the first 18 weeks then every 9 weeks thereafter. Tumor response will be assessed by an IRC and by investigators. Patients who discontinue study drug for reasons other than disease progression (eg, toxicity) will continue to undergo tumor assessments until disease progression, withdrawal of consent, death, or start of a new anticancer therapy, whichever occurs first.

Patients will be evaluated for any AEs occurring up to 30 days after the last dose of study drug or initiation of new anticancer therapy, whichever occurs first (all severity grades, per National Cancer Institute Common Terminology Criteria for Adverse Events [NCI CTCAE] v 4.03) and all immune-related adverse events (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 drug-related SAEs will be recorded by the investigator after treatment discontinuation.

3 STUDY OBJECTIVES

Primary:

To evaluate the efficacy of tislelizumab through Independent Review Committee (IRC) assessed objective response rate (ORR) by Response Evaluation Criteria in Solid Tumors (RECIST) Version (v)1.1 in previously treated, unresectable hepatocellular carcinoma (HCC)

Secondary:

- To assess the efficacy of tislelizumab through duration of response (DOR), progression-free survival (PFS), disease control rate (DCR) and clinical benefit rate (CBR) assessed by IRC and overall survival (OS)
- To assess efficacy of tislelizumab through ORR, DOR, PFS, DCR, and CBR assessed by the investigators
- To assess the safety and tolerability of tislelizumab in patients with previously treated unresectable HCC
- To assess the health-related quality of life (HRQoL) of tislelizumab in patients with previously treated unresectable HCC

Exploratory:

- To assess potential predictive biomarkers
- To characterize the PK of tislelizumab
- To assess host immunogenicity to tislelizumab

4 STUDY ENDPOINTS

4.1 PRIMARY ENDPOINTS

Objective Response Rate (ORR) defined as complete response (CR) or partial response (PR) evaluated by IRC based on RECIST v1.1 in patients with previously treated unresectable HCC.

4.2 SECONDARY ENDPOINTS

- DOR, PFS, DCR and CBR assessed by IRC, and OS
- ORR, DOR. PFS, DCR and CBR assessed by Investigators
- Safety and tolerability assessment of adverse events (AEs), serious adverse events (SAEs), physical examination, vital signs, electrocardiogram (ECG), and laboratory measurements

- HRQoL measured using European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Hepatocellular Carcinoma 18 Questions (EORTC QLQ HCC18) index score, European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30) index-score, and the 5-level version of the European Quality of Life 5-Dimensional Questionnaire (EQ-5D-5L)

4.3 EXPLORATORY ENDPOINTS

- Predictive biomarker (for example, programmed cell death protein ligand-1 [PD-L1] expression and gene expression in tumor tissue)
- PK: Summary of plasma concentrations of tislelizumab
- Immunogenicity: Assessments of immunogenicity of tislelizumab to determine the incidence of anti-drug antibodies (ADAs)

5 SAMPLE SIZE CONSIDERATION

The ORR per IRC is assumed as 15% in this trial. With 228 patients, the power is 0.97 to demonstrate the ORR in patients with previously treated unresectable HCC is statistically higher than the historical rate of 7% in a binomial exact test at one-sided alpha level 0.025. The 95% CI of an observed 15% ORR is (10.6%, 20.3%) when approximately 228 previously treated unresectable HCC patients are enrolled. Within 2nd line or 3rd line plus patients, the 95% CI of the same ORR is (8.6%, 23.5%) when n=100 in each population.

6 STATISTICAL METHODS

6.1 ANALYSIS POPULATIONS

The Safety Analysis Set (SAF) includes all patients who have received one or more doses of tislelizumab. This will be the primary population for efficacy (including HRQoL) and safety analysis.

The Efficacy Evaluable Analysis Set (EFF) includes all patients in the SAF with measurable disease at baseline per RECIST 1.1 who had at least one evaluable post-baseline tumor assessment unless discontinued due to clinical disease progression or death within 7 weeks after the first dose date. This population will only be applied in sensitivity analysis of the primary efficacy endpoint ORR.

The PK Analysis Set (PK) includes all patients in the SAF who contributed at least 1 post-baseline quantifiable PK sample.

6.2 DATA ANALYSIS GENERAL CONSIDERATIONS

Data will be summarized by number of prior systemic anti-cancer therapy (2L vs 3L+) and regions (China and Taiwan vs. Europe).

6.2.1 Definitions and Computations

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

In the situation where the event date is partial or missing, the date will appear partial or missing in the listings; Study day and any corresponding durations will be presented based on the imputation rules specified in Appendix.

Treatment duration: The treatment duration will be calculated as (date of last dose of study treatment – date of first dose of study treatment + 21) for Q3W dose regimens.

Baseline: Unless otherwise specified, a baseline value is defined as the last non-missing value collected before the time of first dose date.

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

6.2.2 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 decimal digit.
- 1 month = 30.4375 days. Number of months is calculated as (days/30.4375) rounded up to 1 decimal digit.
- Age (in years) will be calculated as the integer part of (date of informed consent – date of birth + 1)/365.25
- Missing efficacy or safety data will not be imputed unless otherwise specified.
- For by-visit observed data analyses, percentages will be calculated based on the number of patients with nonmissing data as the denominator, unless otherwise specified.
- For continuous endpoints, summary statistics will include n, mean, standard deviation, median, and range (minimum and maximum).
- For discrete endpoints, summary statistics will include frequencies and percentages.

6.2.3 Handling of Missing Data

Missing data will not be imputed unless otherwise specified elsewhere in the SAP. Missing dates or partially missing dates will be imputed conservatively for adverse events and prior/concomitant medications/procedures. Specific rules for handling of missing or partially missing dates for adverse events and prior/concomitant medications/procedures are provided in Appendix.

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.

6.2.4 Adjustment for Covariates

No adjustments for covariates are planned for primary, secondary and exploratory analyses in the study. Baseline characteristics may be used in the model as covariates as supportive exploratory analyses, if deemed necessary.

6.2.5 Multiplicity Adjustment

No multiplicity adjustments will be made.

6.2.6 Data Integrity

Before pre-specified statistical analysis begins, the integrity of the data should be reviewed to assure 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.

6.3 PATIENT CHARACTERISTICS

6.3.1 Patient Disposition

The number (percentage) of patients with treatment/study ongoing, discontinued from the treatment/study, and reasons for discontinued from treatment/study will be summarized in the SAF.

6.3.2 Protocol Deviations

Protocol deviation criteria will be established; patients with protocol deviations will be identified and documented before the database lock.

Major protocol deviations will be reviewed and summarized. Amongst the major protocol deviations, according to ICH-E3 guidance, those which may impact primary efficacy analysis are identified as critical protocol deviations, and will be further summarized by category such as:

- patients entered the study even though they did not satisfy the entry criteria.
- patients developed withdrawal criteria during the study but were not withdrawn.
- patients received an excluded concomitant treatment.

6.3.3 Demographic and Other Baseline Characteristics

Demographic and other baseline characteristics will be summarized using descriptive statistics in the SAF. Continuous variables include age, BMI (in kg/m^2), body weight (in kg), height (in cm); categorical variables include age group (< 65 vs >= 65 yrs), sex, race, child-bearing potential, tobacco use status, alcohol consumption, ECOG status (0 vs 1).

6.3.4 Disease History and Characteristics

HCC history and characteristics such as BCLC initial staging, time from initial diagnosis to the first study dose, method of initial HCC diagnosis, histological type of cancer , BCLC staging at study entry, time from BCLC staging at study entry to the first study dose, macrovascular invasion, extrahepatic spread, distant metastases, location of metastases and number of metastatic sites will also be summarized.

Alpha-fetoprotein at baseline (descriptively as well as by category >= 200 ng/mL and >= 400 ng/mL) will be summarized as part of disease history.

HCC relevant medical history (Hepatitis B, Hepatitis C, etc.) and ongoing status will be summarized in descriptive statistics.

Total Child-Pugh Score at baseline will be summarized by category (frequency and percentages).

HBV and HCV infection status at baseline based on screening HBV (Hepatitis B Surface Antigen, Hepatitis B Surface Antibody and Hepatitis B Core Antibody) and HCV (Hepatitis C Virus Antibody) test results will be summarized with HBV DNA (Detectable vs. Undetectable) and HCV RNA status (Positive vs. Negative).

The number (percentage) of patients reporting clinically significant symptoms at baseline will be summarized by System Organ Class and MedDRA preferred term. Severity of the symptoms according to NCI-CTCAE will be summarized by maximum grade.

Relevant data listings of disease history will be provided.

6.3.5 Prior Anti-Cancer Treatment

Number of lines of prior systemic anti-cancer therapies will be summarized in frequencies and percentages. Regimen names, best response to the last therapy, duration of sorafenib monotherapy, duration of the last therapy, time from last disease progression to first study dose and reasons of the last therapy discontinued will be summarized.

Total number (percentage) of patients who received any prior liver local regional therapies will be summarized, and the number of types of prior liver local regional therapies will be summarized, and number of patients who received any type of liver local regional therapies will also be summarized by therapy type. Prior liver local regional therapy TACE will be summarized by chemotherapeutic drug names in frequencies and percentages, prior liver local regional radiation therapy (EBRT or SIRT) will also be summarized by radiation therapy type. Treatment intention of the above liver local regional therapies will also be summarized in frequencies and percentages (curative or palliative).

Prior radiotherapies (therapy type, treatment intention, sites etc.) and prior surgeries/procedures (name, intention, and anatomic location) will be summarized similarly. Relevant detail listings will be provided for potential narratives use.

6.3.6 Prior and Concomitant Medications

Prior and concomitant medications will be coded using current version of World Health Organization Drug Dictionary (WHO DD) drug codes 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 name.

Prior medications will be defined as medications that stopped before the first dose of study drug. 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. A listing of prior, concomitant and subsequent medications will be provided.

Concomitant systemic steroid medications and surgeries/procedures used during treatment will be summarized separately.

6.3.7 Medical History

Medical history will be coded using MedDRA codes of the version currently in effect at BeiGene at the time of database lock. The number (percentage) of patients reporting a history of any medical condition (other than HCC relevant history conditions), as recorded on the CRF, will be summarized by system organ class and preferred term. A listing of medical history will be provided.

6.4 EFFICACY ANALYSIS

All efficacy analyses performed on endpoints derived from assessments per Independent Review Committee (IRC) will also be performed on the same endpoints derived from investigators' assessments.

6.4.1 Primary Efficacy Endpoints

Objective Response Rate (ORR)

The number and proportion of patients who achieved confirmed objective tumor response (CR or PR) as evaluated by IRC according to RECIST v1.1 will be summarized. Patients without postbaseline tumor assessment will be considered as non-responders.

Primary analysis of ORR will be based on SAF. Analysis of ORR will also be conducted in the EFF as a sensitivity analysis.

A statistical testing of comparison of ORR of study drug over a pre-specified historical ORR (7%) in null hypothesis will be performed using Binomial exact method, demonstration of superiority of test drug will be claimed if one-sided p-value is < 0.025.

A two-sided binomial exact 95% confidence interval (CI) of ORR will also be constructed to assess the precision of the point estimate of ORR.

The primary efficacy analysis will be conducted when mature response rate data have been observed, estimated approximately 6 months after the last patient in (LPI). Additional efficacy analysis will be conducted approximately 12 months after LPI.

6.4.2 Secondary Efficacy Endpoints

Duration of Response (DOR)

Duration of response for responders (CR or PR) is defined as the time interval between the date of the earliest qualifying response and the date of PD or death for any cause. Duration of response analysis will only include responders. Censoring rule for DOR will follow PFS censoring rule (described below).

Kaplan-Meier method will be used to estimate median time and 95% confidence interval for duration of response.

Progression Free Survival (PFS)

PFS is defined as the time from the date of first dose of study treatment to disease progression or death, whichever is earlier. Kaplan Meier method along with the corresponding 95% CI based on Greenwood's formula will be used to construct the PFS time point estimates which are defined as the percentages of patients in the analysis population who remain alive and progression-free at the specified time points (ie, 3 or 6 month). Kaplan Meier curves will be constructed to provide a visual description of the PFS rate versus time.

The PFS derivation rules in this SAP follow Food and Drug Administration (FDA) "Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics (2007)" with some modifications.

Table 1 shows the primary censoring rules for the derivation of PFS using RECIST v1.1 criteria based upon radiological tumor assessment.

Table 1Censoring Rules for Analysis of Progression-Free Survival Per RECIST v1.1

No.	Situation	Date of Progression or Censoring	Outcome
1	No baseline tumor assessments	Date of the first dose	Censored
2	Progression documented on scheduled visit or 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	Death before first PD assessment	Date of death	Progressed
6	Death between adequate assessment visits*	Date of death	Progressed
7	Death or progression after more than one missed visit**	Date of last adequate radiologic assessment prior to missing occurred	Censored

Abbreviations: CR = complete response, PD = progressive disease, PR = partial response, SD = stable disease,

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

** More than one missed visit is defined as the duration between the last tumor assessment and death or PD is longer than 13 weeks week in the first 18 weeks, and longer than 19 weeks thereafter.

The priority of the censoring rules in the primary analysis is as follows:

1. If the patient had an event of PD or death, the following sequence will be applied:

    a) If a patient did not have baseline tumor assessment (No. 1), the patient will be censored on date of the first dose of study treatment. However, if the patient died within 91 days (13 weeks) after the first dose date and did not receive new anticancer treatment, the date of death will be the PFS event date (not censored).

    b) If a patient had new anticancer treatment before the earliest event of PD or death (No. 4), the patient will be censored on the date of the last tumor assessment prior to or on the date of new anticancer treatment.

    c) If a patient had more than one missed tumor assessment visits before the earliest event of PD or death (No. 7), the patient will be censored on the date of the last adequate radiologic assessment prior to missing tumor assessment.

    d) Otherwise, the earliest event date will be used (No. 2, No. 5, or No. 6).

2. If a patient did not have PD or death, the censoring date will be the earliest censoring date if the patient met multiple censoring criteria (No. 1, No. 3, No. 4).

In addition, as a sensitivity analysis, clinical PD will be considered as an event in calculation of PFS.

Disease Control Rate (DCR)

DCR is defined as the proportion of patients whose best overall response (BOR) is CR, PR or SD by IRC and Investigator per RECIST v1.1. Patients without postbaseline tumor assessment will be considered as failure in DCR.

A two-sided binomial exact 95% confidence interval (CI) of DCR will also be constructed to assess the precision of the point estimate of DCR.

Clinical Benefit Rate (CBR)

CBR is defined as the proportion of patients who have CR, PR, or SD of ≥ 24 weeks in duration, assessed by IRC and Investigator per RECIST v1.1. SD duration is measured from the start of study treatment until the time that the criteria for progression are met.

A two-sided binomial exact 95% confidence interval (CI) of CBR will also be constructed to assess the precision of the point estimate of CBR.

Overall Survival (OS)

Overall survival is defined as the time from the date of the first dose of study treatment to death. Patients who remained alive at data cutoff or discontinuation of the study (discontinued study due to reasons other than "Death") will be censored at the time of the last date the patient was known to be alive.

The OS time point estimates, defined as the percentages of patients in the analysis population who remain alive at the specified time points (ie, 3, 6 months etc.), will be estimated using the Kaplan-Meier method along with the corresponding 95% CI constructed using Greenwood's formula. The Kaplan-Meier estimates of OS will be plotted over time.

Waterfall plots will be provided for the maximum tumor shrinkage based on target lesion. The maximum tumor shrinkage based on target lesion used in the plots will be listed.

In addition, number of patients who received post-treatment anti-cancer therapies (surgery/procedure, radiation, or systemic therapies) and regimen will be summarized.

6.4.3 Subgroup Analyses

Subgroup analysis on key efficacy endpoints (ORR, OS etc.) will be conducted to explore the consistence of efficacy across variety of subgroups. Subgroup variables may include, but not limited to:

- age (≤ 65 vs > 65 years)
- gender (Female vs Male)
- macrovascular invasion and/or extrahepatic spread (present or absent)
- macrovascular invasion (present or absent)
- extrahepatic spread (present or absent)
- hepatitis viral status (HBV vs HCV vs Uninfected)
- ECOG PS (0 vs 1)
- Screening alpha-fetoprotein (<= 200 ng/mL vs > 200 ng/mL and <= 400 ng/mL vs > 400 ng/mL)
- BCLC stage at study entry (Stage B vs Stage C)
- prior loco-regional therapy (yes or no)

- # of prior TACE loco-regional therapies
- PD-L1 expression

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

6.5 SAFETY ANALYSES

All safety analyses will be performed in the SAF. Treatment extent of exposure, incidence of treatment emergent adverse events, death, laboratory, vital signs, ECOG Performance Status and ECG will be analyzed; Abnormal values will be flagged.

6.5.1 Extent of Exposure

Descriptive Statistics (n, mean, standard deviation, median, minimum and maximum) will be estimated for continuous variables including total dose in mg, duration of treatment, actual dose intensity, relative dose intensity, number of cycles. Categorical variables include duration of treatment in months, number of cycles received, number of patients with dose modification (dose delay or interruption).

In addition, frequency of dose delay and frequency of dose interruption will be summarized by categories (0, 1, ≥2).

Actual dose intensity (ADI) per patient (mg/cycle) which is defined as sum of actual dose received divided by duration of treatment and relative dose intensity (RDI) per patient (%) defined as actual dose intensity/planned dose intensity per patient will be summarized.

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

6.5.2 Adverse Events

AEs will be graded by the investigators using CTCAE v4.03. The AE verbatim descriptions (investigator terms from the CRF) will be classified into standardized medical terminology using the Medical Dictionary for Regulatory Activities (MedDRA). Adverse events will be coded to the MedDRA lower level term closest to the verbatim term. The linked MedDRA preferred term (PT) and primary system organ class (SOC) are also captured in the database.

A treatment emergent adverse event (TEAE) is defined as an AE that had an onset date on or after the date of the first dose of study treatment through 30 days after the last dose (permanent discontinuation of study treatment) or initiation of new anti-cancer therapy whichever occurs earlier. The TEAE classification also includes immune-related AEs and related serious AEs that are recorded up to 90 days after discontinuation from treatment. 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.

An overview table, including the incidence of and the number of patients with TEAEs, treatment-emergent serious adverse events (SAEs), treatment-related TEAEs, TEAEs with grade 3 or above, treatment-related SAEs, TEAEs that led to death, TEAEs that led to treatment discontinuation,

TEAEs that led to dose hold/interruption, potential immune-related TEAEs and infusion related reactions will be provided. Treatment-related AEs include those events considered by the investigator to be related to study treatment or with missing assessment of the causal relationship.

The incidence of TEAEs will be reported as the number (percentage) of patients with TEAEs by SOC and PT. A patient will be counted only once by the highest severity grade according to CTCAE v.4.03 within an SOC and PT, even if the patient experienced more than 1 TEAE within a specific SOC and PT.

The number (percentage) of patients with treatment-emergent SAEs, treatment-related TEAEs, TEAEs with grade 3 or above, treatment-related SAEs, TEAEs that led to death, TEAEs that led to treatment discontinuation, TEAEs that led to dose hold/interruption, potential immune-related TEAEs, and infusion related reactions (IRR) will be summarized by SOC and PT.

1. Overall summary of TEAEs
2. All TEAEs by SOC and PT
3. Treatment-related TEAEs by SOC and PT
4. All TEAEs with Grade 3 or above by SOC and PT
5. TEAEs by maximum severity
6. Treatment-related Grade 3 or above TEAEs by SOC and PT
7. All serious TEAEs by SOC and PT
8. Treatment-related serious TEAEs by SOC and PT
9. TEAEs leading to study drug permanently discontinued
10. Treatment-related TEAEs leading to study drug permanently discontinued
11. TEAEs leading to study drug dose held/interrupted
12. Treatment-related TEAEs leading to study drug dose hold/interrupted
13. TEAEs leading to death
14. Treatment-related TEAEs leading to death
15. immune-related TEAEs (irTEAEs)
16. TEAEs considered as infusion related reactions (IRR)

Patient data listings of all AEs, SAEs, AEs that led to death and AEs that led to treatment discontinuation will be provided.

All deaths and causes of death will be reported, including those which occurred during the study treatment period and those reported during the survival follow-up period after treatment completion/discontinuation.

6.5.3 Laboratory Values

Laboratory results for hematology, biochemistry and coagulation will be summarized using System International (SI) units, as appropriate. For all quantitative parameters listed in Table 2, the actual value and the change from baseline to each post-baseline visit and to the end of treatment will be summarized by visit using descriptive statistics (n, mean, standard deviation, median, minimum, maximum for continuous variables; n [%] for categorical variables). Unscheduled measurements will not be summarized by visit but will be included in calculation of minimum/maximum change from baseline.

Laboratory parameters (Alkaline phosphatase, Alanine aminotransferase, Aspartate etc.) that are graded in NCI CTCAE (v. 4.03) will be summarized by shifts from baseline CTCAE grades to worst post-baseline grades. In the summary of laboratory parameters by CTCAE grade, parameters with CTCAE grading in both high and low directions will be summarized separately.

Table2 Clinical Laboratory Assessments

Serum Chemistry	Hematology	Coagulation
Alkaline phosphatase Alanine aminotransferase Aspartate aminotransferase Albumin Direct bilirubin Total bilirubin Blood urea nitrogen or urea Creatinine Glucose Lactate dehydrogenase Total protein Potassium Sodium Creatine kinase (CK) Creatine kinase-cardiac muscle isoenzyme (CK-MB)[2]	Hematocrit Hemoglobin Platelet count WBC count Neutrophil count Lymphocyte count	Prothrombin time Activated Partial Thromboplastin Time International Normalized Ratio

Furthermore, to evaluate liver injury, potential Hy's Law cases, i.e. patients with any ALT or AST > 3xULN, patients with any TBL > 2x ULN, patients who met both criteria above, and patients who met both above criteria with ALP < 2xULN observed will be summarized.

In addition, change from baseline in alpha-fetoprotein (AFP) will be summarized.

6.5.4 Vital Signs

The change from baseline vital signs in blood pressure, pulse rate, weight, temperature in Celsius will be summarized in descriptive statistics. Potentially Clinically Significant (PCS) vital signs such as pulse rate (bpm) <= 45, >= 120, systolic BP (mmHg) <= 90, >= 160, and diastolic BP (mmHg) <=60, >= 100 will be summarized.

6.5.5 Physical Examination

Physical examination will be assessed during screening and study visits. Physical examination findings prior to first dose of study treatment will be collected in medical history, clinically significant abnormalities found in physical examination will be reported in adverse events. No separate physical examination data will be collected and reported in this study.

6.5.6 Ophthalmologic Examination

A data listing of ophthalmologic examination results will be provided.

6.5.7 Electrocardiograms (ECG)

ECG will be performed at the baseline and multiple time points (refer the time points to the protocol study assessments and procedures schedule) after the start of treatment. Observed and change from baseline in ECG will be summarized descriptively over time by visit. Observed post-baseline QTc interval measurement and increase from baseline by pre-specified thresholds will be summarized by category.

6.5.8 Eastern Cooperative Oncology Group (ECOG)

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

6.6 PHARMACOKINETIC ANALYSES

Pharmacokinetic samples will be collected in this study (refer to Appendix 1 of protocol), and only from patients receiving tislelizumab at 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 (C_{trough}) will be tabulated and summarized by visit/cycle at which these concentrations were collected. Descriptive statistics will include means, geometric mean, medians, CV, ranges, and standard deviations, as appropriate.

6.7 IMMUNOGENICITY ANALYSES

Human anti-drug antibodies (ADA) to tislelizumab will be assessed during the study as defined in the protocol.

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.
- **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 in the last time point.

- **Transient ADA response** is defined as treatment-induced response that is not considered persistent.
- **Neutralizing ADA** is defined as ADA that inhibits or reduces the pharmacological activity.

ADA response endpoints:

- **ADA incidence** is defined as sum of treatment-emergent ADA, which include both 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.

6.8 BIOMARKER ANALYSES

The primary predictive biomarker analysis is based on a subset of the patients with both a valid PD-L1 expression and/or tumor-infiltrating lymphocytes (TILs) measurement and at least one disease assessment post-treatment. A supportive analysis is based on patients with a valid PD-L1 expression and/or TILs measurement, irrespective of the availability of post-treatment disease assessments. Exploratory analyses of other candidate predictive biomarkers, including but not limited to gene expression profiling, will be conducted similarly.

6.9 ANALYSES OF HEALTH-RELATED QUALITY OF LIFE

Descriptive summary statistics (n, mean, standard deviation, median, minimum, maximum) of Health-Related Quality of Life (HRQoL) measured using European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Hepatocellular Carcinoma 18 Questions (EORTC QLQ HCC18) index score, European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30) index-score, and the 5-level version of the European Quality of Life 5-Dimensional Questionnaire (EQ-5D-5L) will be applied in this study.

HRQoL scores, obtained from the generic quality of life of cancer patients (EORTC QLQ-C30), HCC specific module (EORTC QLQ-HCC18) and EQ-5D-5L will be analyzed in the Safety Population. There are 1 global health scale (Aaronson NK, et al., 1993; Fayers PM, et al., 2001), five functional scales (physical, role, cognitive, emotional, and social), three symptom scales (fatigue, pain, and nausea and vomiting), and six single items (dyspnoea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties in EORTC QLQ-C30 and 6 scales and 2 single items in EORTC QLQ-HCC18 (see Appendix 10.2).

The most relevant scales are EORTC-C30 and HCC18 index scores, 4 subscales in QLQ-C30 (physical function, role function, pain, diarrhea scales) and 2 subscales in the HCC disease specific module (nutrition and body image scale). They have been selected as more important disease

specific symptoms. The EORTC-HCC18 symptom scores can also provide patient-reported indices of adverse effects of treatment.

For the EORTC QLQ-C30 and EORTC QLQ HCC18 at each visit, raw score for functional scales and symptom scales will be calculated based on questionnaire items. Raw scores for the functional scale/symptom scale/single items will be transformed into 0-100 scale via linear transformation. For details, refer to the Appendix 10.2.

The EQ-5D-5L comprises a descriptive system and an EQ Visual Analogue scale (EQ VAS) with following 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems, and extreme problems. The EQ VAS records the respondent's self-rated health on a 0 to 100 scale, with 100 labelled 'the best health you can imagine' and 0 'the worst health you can imagine'.

Analysis Method

There is no multiplicity adjustment for the HRQoL analysis in this single arm study. Descriptive statistics will be used for all analyses. In addition to index scores, EORTC QLQ-C30 physical function, role function, pain, diarrhea scales and EORTC QLQ HCC-18 nutrition and body image scale have been selected as the more important parameters among all scales and items based on clinical relevance and hypothesized association with PD1 treatment. While the remaining scales and items will be summarized similarly.

Completion rates

Completion rates for the EORTC QLQ-C30, QLQ-HCC18 and EQ-5D-L modules in the Safety Population will be summarized separately at each visit. A Questionnaire module is considered complete if all questions are answered. In addition, the completion rate in patients still on treatment (whose discontinuation date is at least 7 days after the visit) will also be summarized in a tabular form.

For those who are still in treatment, the reasons of missing HRQoL will be documented in the datasets.

Change from Baseline by Visit

For each of the scale, single item and index score, summary statistics at each assessment time point and change from baseline will be presented in tables.

Time to Deterioration (Clinically Meaningful Worsening of HRQoL)

Time to deterioration (TTD, i.e. time to clinical meaningful worsening) is defined as the time from the first dose of tislelizumab to first onset time at which symptom declined by an identified clinically important threshold in terms of increase of index score. A 10 points difference in QLQ-

C30 is generally considered clinically significant (Aaronson et al, 1993). Since QLQ-HCC18 was developed in a similar approach, the same 10 points difference will be used in QLQ-HCC18. The threshold of 10-point is consistent with the results from an international HCC18 validation study (Chie 2012) in which the average standard deviations of HCC 18 scales at baseline in 5 HCC clinical groups are less than 20. Since a change of half a standard deviation of the baseline score is considered as clinically significant (Frans 2014), it is reasonable to use 10-point as clinical important threshold. Patients without deterioration (Note: deterioration means increase in index score) will be censored at the last date at which the specific scale score is available, including those who discontinued treatment due to any reasons (e.g. lost to follow-up, disease progression and death).

The median time to clinical meaningful worsening of C30 and HCC18 will be calculated using Kaplan-Meier estimates, and presented with 2-sided 95% CIs.

EQ-5D-5L

Five level response to EQ-5D-5L will be summarized as a categorical variable by tabulating frequency of each response category by visit. The self-rated health state scale at each visit and its change from baseline will be summarized in descriptive statistics (n, mean, standard deviation, median, minimum, maximum).

7 INTERIM ANALYSIS

No interim analysis is planned.

8 CHANGES IN THE PLANNED ANALYSIS

Not applicable

9 REFERENCES

Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organisation for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. Journal National Cancer Institute. 1993; 85:365-76.

Brookmeyer R, Crowley J. A confidence interval for the median survival time, Biometrics 1982; 38(1): 29-41.

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 whit primary liver cancer, the European Organization for Research and Treatment of Cancer Quality-of-Life Questionnaire HCC18. Hepatology v55, 1122- 1129, 2012

Common Toxicity Criteria Version 4.03. Cancer Therapy Evaluation Program. June 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.

Food and Drug Administration Center for Drug Evaluation Research CDER and Center for Biologics Evaluation and Research (2007). FDA Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics (2007).

Frans FA, Nieuwkrek, PT, Met R, Statistical or clinical improvement determining the minimally important difference for the vascular quality of life questionnaire in patients with critical limb ischemia. European journal of vascular and endovascular surgery 2014 v47 180-186.

Greenwood M. “The natural duration of cancer”. Reports on Public Health and Medical Patients (London: Her Majesty’s Stationery Office). 1926; 33:1-26.

International Conference on Harmonization Note for Guidance on Good Clinical Practice (CPMP/ICH/135/95). July 1996.

Kaplan E, Meier P. Nonparametric estimation from incomplete observations. Journal of American Statistical Association. 1958; 53:457-81.

Li L, Mo F, 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.

Shankar, G., Arkin, S., Cocea, L. et al. Assessment and Reporting of the Clinical Immunogenicity of Therapeutic Proteins and Peptides—Harmonized Terminology and Tactical Recommendations. AAPS J (2014) 16: 658.

10 APPENDIX

10.1 IMPUTATION OF MISSING/PARTIALLY MISSING DATES

Missing data will not be imputed unless otherwise specified. The imputation rule for the safety analyses will be used to address the issues with partial dates.

i) When the start date or end date of an adverse event is partially missing, the date will be imputed to determine whether the adverse event is treatment-emergent. When in doubt, the adverse event will be considered treatment emergent by default. The following rules will be applied to impute partial dates for adverse events:

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 year of the start date is missing, or start date is completely missing, do not impute.

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 year of the end date is missing, end date is completely missing, do not impute

ii) When the start date or end date of a medical or disease history/medication/therapy (systemic or radiotherapy)/surgery or procedure is partially missing, the date will be imputed to determine whether the medical or disease history/medication/therapy (systemic or radiotherapy)/surgery or procedure is prior or concomitant or post. The following rules will be applied to impute partial dates:

If the start date 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 end date 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 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 the date is known to be prior study treatment, e.g. for medical history, disease history, or prior therapies etc., the imputed date should be compared against the first dose date. If it is after the first dose date, then replace the imputed date with the first dose date – 1.

If the date is known to be post study treatment, e.g. for post-treatment subsequent therapies, the imputed date should be compared against the last dose date. If it is before the last dose date, then replace the imputed date with the last dose date + 1.

10.2 HEALTH RELATED QUALITY OF LIFE

HRQoL scores, obtained from the generic quality of life of cancer patients (EORTC QLQ-C30), HCC specific module (EORTC QLQ-HCC18) and EQ-5D-5L will be analyzed in the Safety Population. There are 1 global health status, 8 scales and 6 single items in EORTC QLQ-C30 and 6 scales and 2 single items in EORTC QLQ-HCC18. The EORTC QLQ-C30 and HCC18 index-scores (Li, 2017) and scale scores will be calculated as described below.

Scale Derivation

Scoring Process
The principle for scoring applies to all scales/scores: Raw scores are calculated as the average of the items that contribute to the scale.

A linear transformation to standardize the raw scores is utilized, so that the scores are ranged from 0 to 100. Increases in scores for functional domains (e.g., physical, role, social, emotional, etc.) are improvements while increases in scores for symptoms (e.g., fatigue, vomiting and nausea, diarrhea, pain, etc.) are deteriorations.

Missing Items
In this trial, an ePRO system will require patients to complete all Patient-Repot Outcome (PRO) questions, although it can be interrupted, in which case the form will save as “incomplete” and allow the patient to choose to either re-start the form or complete the remainder of it.
In case of an “incomplete” form in the analysis, all the completed items could be used to calculate the score if at least half of the items for a scale are answered. Otherwise, the scale score is set to missing.

Raw Score
For all scores, the raw score (RS), is the mean of the component items
RS=(l1+l2+…+ln)/n

Derived Scale
The derived scales are obtained from the raw scores as defined in the EORTC manual. The derived scales have a more intuitive interpretation: larger function scale or global health status /

QoL are improvements while larger symptom scales (e.g., pain, nausea, etc.) are deteriorations.

The derivation formulas are as follows.

Derived function scale, symptom scale and index scale

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

HCC18 index-score=∑ (scores of Fatigue, Body image, Jaundice, Nutrition, Pain, Fever, Sex life, Abdominal distension) ÷ 8

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