Official Title of Study:

A Randomized, Multi-center Phase III Study of Nivolumab versus Sorafenib as First-Line Treatment in Patients with Advanced Hepatocellular Carcinoma

NCT Number: NCT02576509

Document Date (Date in which document was last revised): July 2, 2018

STATISTICAL ANALYSIS PLAN

A RANDOMIZED, MULTI-CENTER PHASE III STUDY OF NIVOLUMAB VERSUS SORAFENIB AS FIRST-LINE TREATMENT IN PATIENTS WITH ADVANCED HEPATOCELLULAR CARCINOMA

(CHECKMATE 459: CHECKPOINT PATHWAY AND NIVOLUMAB CLINICAL TRIAL EVALUATION 459)

PROTOCOL(S) CA209459

VERSION # 3.0

TABLE OF CONTENTS

[redacted]

[redacted]

[redacted]

2 STUDY DESCRIPTION

2.1 Study Design

This is an open-label, two-arm, randomized, Phase 3 study in adult (≥ 18 years) male and female subjects with advanced HCC. Subjects must not be amenable for management with surgery or loco-regional therapy or have progressed after surgery or loco-regional therapy. Subjects must not have received prior systemic therapy for advanced HCC in keeping with the first-line setting of this study.

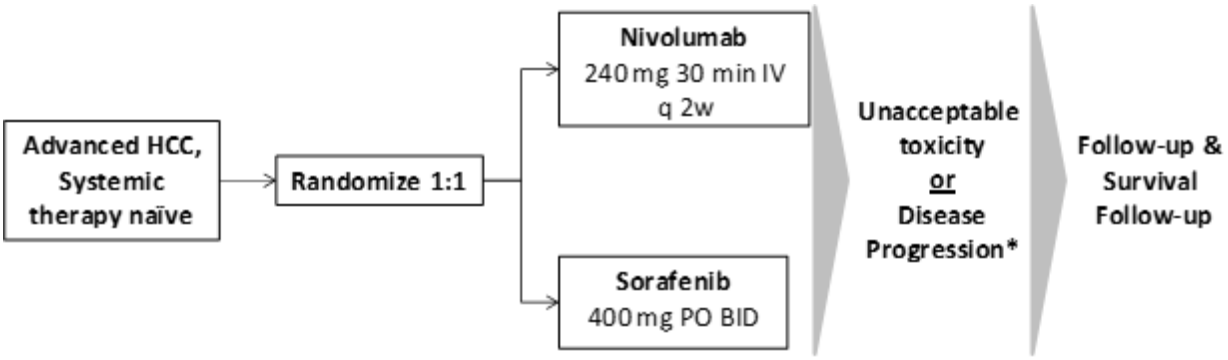
After screening for eligibility and signing of informed consent, qualified subjects will be randomized in a 1:1 ratio to receive nivolumab or sorafenib. It is expected that approximately 908 subjects will be screened for the study with approximately 726 subjects will be randomized, e.g., approximately 363 subjects will be randomized to each study arm. Randomization will be stratified by etiology (HCV vs. non-HCV [ie, HBV and HCC with no history of hepatitis virus infection]), vascular invasion and/or extrahepatic spread (present or absent), geography (Asia vs Non-Asia).

Subjects will receive open-label treatment with one of the following:

- Nivolumab 240 mg IV every 2 weeks until disease progression or unacceptable toxicity
- Sorafenib 400 mg PO BID until disease progression or unacceptable toxicity

The study design schematic is presented in Figure 2.1-1.

Figure 2.1-1: Study Design Schematic



* Subjects may be treated beyond disease progression under protocol-defined conditions.

Stratification factors:
- Etiology (HCV vs. non-HCV [ie, HBV- and HCC with no history of hepatitis virus infection])
- Vascular invasion &/or extrahepatic spread (present or absent),
- Geography (Asia vs Non-Asia).

On treatment visits will occur at Day 1 and every 2 weeks thereafter. Tumor imaging assessments will occur 8 weeks from the date of randomization (+/-1 wk), then every 8 weeks (+/- 1 wk) thereafter up to 48 weeks, then it will be every 12 weeks (+/- 1 week) until disease progression or treatment is discontinued (whichever occurs later). Subjects will be treated until unacceptable toxicity or disease progression. Following discontinuation of therapy, safety will be assessed through post-treatment Follow Up visit 2 (~ 100 days from last dose). Survival status will be assessed every 3 months after follow up visits are completed, and may be completed via telephone or in person visits.

On both arms, treatment beyond initial investigator-assessed RECIST 1.1 defined progression is permitted if the subject has investigator assessed clinical benefit and is tolerating study drug, as specified in the protocol[31].

This study's primary endpoint is OS in all randomized subjects. Its secondary endpoints include objective response rate (ORR) and progression-free survival (PFS) based on tumor response assessed by a blinded and independent central review (BICR) according to RECIST 1.1.

2.2 Treatment Assignment

CA209459 is a randomized, open-label study. After the subject's initial eligibility is established and informed consent has been obtained, the subject must be enrolled into the study by calling an interactive voice response system (IVRS) to obtain the subject number. Every subject that signs the informed consent form must be assigned a subject number in IVRS. Specific instructions for using IVRS will be provided to the investigational site in a separate document.

The investigator or designee will register the subject for enrollment by following the enrollment procedures established by BMS. The following information is required for enrollment:

- Date that informed consent was obtained

- Date of birth
- Gender at birth
- Viral status at time of enrollment, if known

Once enrolled in IVRS, enrolled subjects that have met all eligibility criteria will be ready to be randomized through IVRS. The following information is required for site entry in the IVRS prior to subject randomization:

- Subject number
- Date of birth
- Vascular invasion and/or extrahepatic spread (present or absent)

Additional information will be automatically transferred to the IVRS by the Central Lab and must be available prior to subject randomization:

- Etiology (HCV vs non-HCV)
- Confirmation of tumor tissue sample receipt and acceptability at Central Laboratory

Subjects meeting all eligibility criteria will be randomized in a 1:1 ratio to nivolumab or sorafenib, stratified by the following factors:

- Etiology (HCV vs non-HCV)
- Vascular invasion and/or extrahepatic spread (present or absent)
- Geography (Asia vs Non-Asia)

The randomization procedures will be carried out via permuted blocks within each stratum. The exact procedures for using the IVRS will be detailed in the IVRS manual.

2.3 Blinding and Unblinding

This is an open-label study.

2.4 Protocol Amendments

This statistical analysis plan (SAP) incorporates the amendments in Table 2.4-1.

Table 2.4-1: Protocol Amendments

Amendment	Date of Issue	Summary of Major Changes
Amendment 16	15-Aug-2017	• Changed ORR from a co-primary objective to a secondary objective of the study. • OS is the sole primary endpoint of the study.
Revised Protocol 02	24-Aug-2016	Incorporates Amendment 11
Amendment 11	24-Aug-2016	• Updated Study Director/Medical Monitor information • Primary endpoint and objective changed from time

Table 2.4-1: Protocol Amendments

Amendment	Date of Issue	Summary of Major Changes
		to progression (TTP) to objective response rate (ORR) – Added requirement for confirmatory scan to be performed for CR/PR assessment of best overall response (BOR) • Management algorithms updated per revised nivolumab IB
Revised Protocol 01	01-Oct-2015	Incorporates Amendment 02
Amendment 02	01-Oct-2015	Addition of collection of peripheral blood mononuclear cells (PBMCs) and myeloid derived suppressor cells (MDSCs) collected from subjects at baseline from selected sites, reduction of frequency of HCV RNA testing for HCV infected subjects, clarification of locoregional therapy inclusion criteria, as well as other miscellaneous clarifications and corrections.

2.5 Data Monitoring Committee

A DMC will be established to provide oversight and safety and efficacy considerations in protocol CA209459. The DMC will provide advice to the sponsor regarding actions the committee deems necessary for the continuing protection of subjects enrolled in the study. The DMC will be charged with assessing such actions in light of an acceptable benefit/risk profile for nivolumab. The DMC will act in an advisory capacity to BMS and will monitor subject safety and evaluate the available efficacy data for the study.

The BMS clinical study leadership will have responsibility for overall conduct of the study including managing the communication of study data. The group will be responsible for promptly reviewing the DMC recommendations, for providing guidance regarding the continuation or termination of the study, and for determining whether amendments to the protocol or changes to the study conduct are required. Adjudicated events will be submitted to the DMC and Health Authorities when required for review on a specified timeframe in accordance with the adjudication documentation.

Details of the DMC responsibilities and procedures will be specified in the DMC charter[30]. Details of the statistical analysis plan supporting DMC review will be specified in a DMC SAP.

2.6 Blinded Radiology Review Committee

In addition to local tumor assessments, images from this study will undergo a blinded, independent central review (BICR) to assess response based on the RECIST 1.1 and mRECIST assessment criteria. The centrally reviewed response data will be used in the analyses of ORR, PFS, duration of response (DOR), and time to response (TTR). All final determinations on centrally reviewed image-based endpoints will be made based on the independent assessments

for a uniform and unbiased assessment of outcome. Details of the procedures and the criteria for the central review are defined in a separate imaging charter[32].

3 OBJECTIVES

3.1 Primary

To compare OS of nivolumab to sorafenib in subjects with advanced HCC who have not received prior systemic therapy.

3.2 Secondary

- To compare ORR of nivolumab to sorafenib. ORR will be determined from assessment by BICR based on RECIST 1.1
- To compare PFS of nivolumab and sorafenib. PFS will be determined from assessments by BICR based on RECIST 1.1
- To evaluate the relationship between tumor PD-L1 expression and efficacy

4 ENDPOINTS

4.1 General Definitions of Efficacy Endpoints based on Tumor Assessment

Tumor assessment based efficacy endpoints such as ORR and PFS will be determined per RECIST 1.1 or mRECIST for HCC by BICR or investigators in different populations of interest. The following criteria will be used and specified in definitions of related endpoints in Section 4.2, 4.3, and 4.4: BICR per RECIST 1.1, investigator per RECIST 1.1, and BICR per mRECIST for HCC.

4.1.1 ORR

ORR is defined as the proportion of subjects whose best overall response (BOR) is either a complete response (CR) or partial response (PR) among all randomized subjects in a population of interest. BOR is defined as the best response designation recorded between the date of randomization and the date of first objectively documented progression or the date of subsequent anti-cancer therapy, whichever occurs first. For subjects without documented progression or subsequent anti-cancer therapy, all available response designations will contribute to the BOR determination. For a BOR of CR or PR, the initial response assessment must be confirmed by a consecutive assessment no less than 4 weeks (28 days) later. In the case of stable disease (SD), measurements must have met the SD criteria at least once after randomization at a minimum of 7 weeks (49 days). **Responder** is defined as a subject whose BOR is either a CR or PR.

Tumor imaging assessments will occur 8 weeks from the date of randomization (+/-1 wk), then every 8 weeks (+/- 1 wk) thereafter up to 48 weeks, then it will be every 12 weeks (+/- 1 week) until disease progression or treatment is discontinued (whichever occurs later).

4.1.2 PFS

PFS is defined as the time from the date of randomization to the date of the first objectively documented tumor progression or death due to any cause in all randomized subjects. Subjects who die without a reported prior progression and without initiation of subsequent anti-cancer therapy will be considered to have progressed on the date of their death. Subjects who did not progress or die will be censored on the date of their last tumor assessment prior to subsequent anti-cancer therapy (if there is any). Subjects who did not have baseline tumor assessment will be censored on the date they were randomized. Subjects who did not have any on study tumor assessments and did not die will be censored on the date they were randomized. Censoring rules for the primary analysis of PFS are summarized in Table 4.1.2-1. Further explanation for various censoring scenarios for the primary definition of PFS are presented in Figure 4.1.2-1.

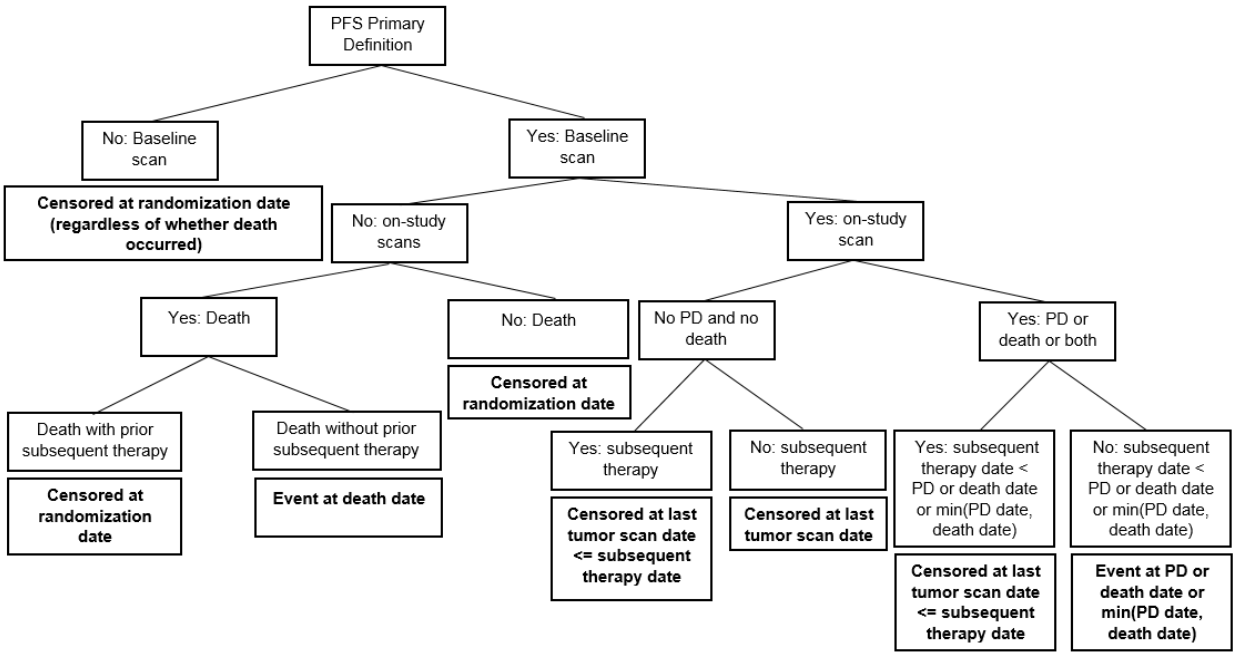
Table 4.1.2-1: Censoring Scheme used in Primary Analysis of PFS

Situation	Date of Progression or Censoring	Outcome
No baseline tumor assessment	Date of randomization	Censored
No on study tumor assessments and no death	Date of randomization	Censored

Table 4.1.2-1: Censoring Scheme used in Primary Analysis of PFS

Situation	Date of Progression or Censoring	Outcome
Documented progression	Date of the first documented tumor progression	Progressed
No progression and no death	Date of last tumor assessment	Censored
Subsequent anti-cancer therapies started without a prior reported progression	Date of last tumor assessment prior to initiation of the subsequent therapy	Censored
Death without progression and without initiation of subsequent anti-cancer therapy	Date of death	Progressed

Figure 4.1.2-1: Graphical display of PFS Primary Definition

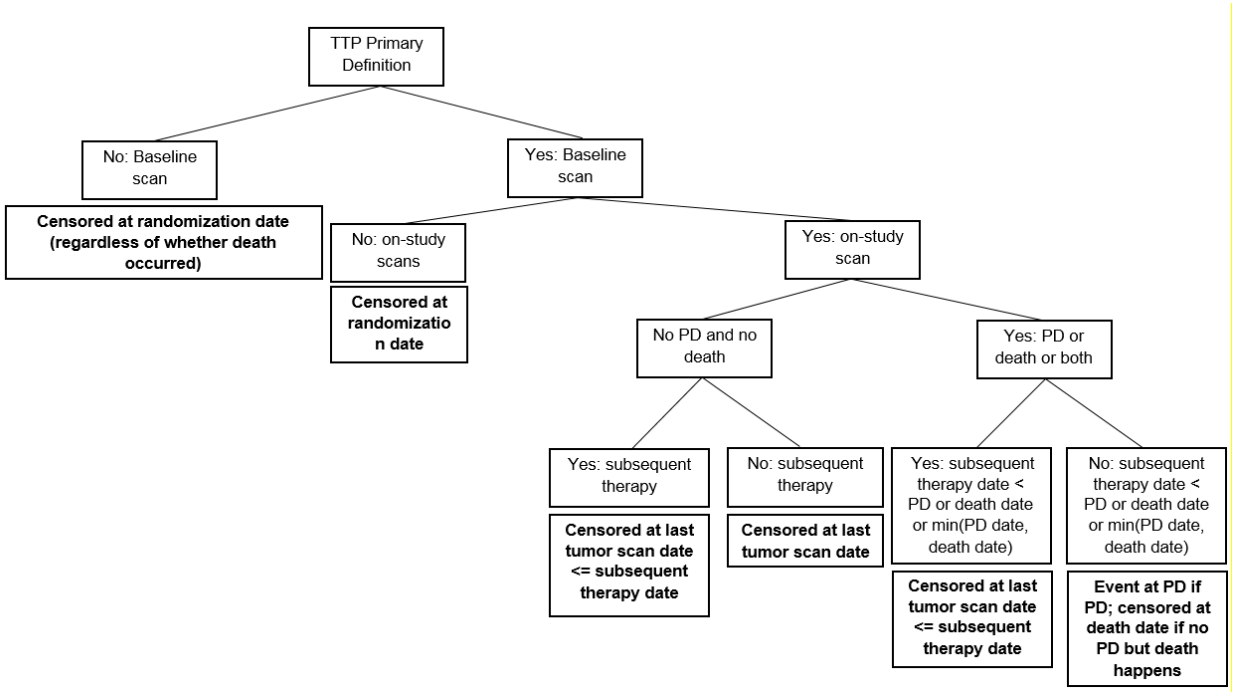


4.1.3 TTP

Time to Progression (TTP) is defined as the time from the date of randomization to the date of the first objectively documented tumor progression in all randomized subjects. The censoring scheme for the primary definition of TTP follows the graphical display in

Figure 4.1.2-2.

Figure 4.1.2-2: Graphical display of TTP Primary Definition



4.1.4 DOR

DOR is defined as the time between the date of first documented response (CR or PR as defined in Section 4.1.1) to the date of first documented tumor progression or death due to any cause, whichever occurs first. Subjects who neither progress nor die will be censored on the date of their last evaluable tumor assessment. Subjects who started any subsequent anti-cancer therapy without a prior reported progression will be censored at the last evaluable tumor assessment prior to or on the date of initiation of the subsequent anti-cancer therapy. Essentially, the DOR will be censored according to the same censoring scheme as the one for the primary definition of PFS (Section 4.1.2, with baseline and on-study scans available). DOR will be evaluated for responders (i.e. subjects with confirmed CR or PR) only.

4.1.5 TTR

TTR is defined as the time from randomization to the date of the first confirmed CR or PR. TTR is derived for responders only.

4.1.6 DCR

DCR is defined as the proportion of subjects whose BOR is CR or PR or SD or NON-CR/NON-PD among all randomized subjects of a population of interest **based on BICR-assessed** BOR. **Based on investigator-assessed BOR**, DCR is defined as the proportion of subjects whose BOR is CR or PR or SD among all randomized subjects of a population of interest.

4.1.7 DDC

DDC is defined as the time from the date of randomization to the date of the first documented tumor progression or death due to any cause in subjects whose BOR is CR or PR or SD or NON-CR/NON-PD **based on BICR-assessed BOR**. Disease of these subjects is considered to be controlled till progression or death. DDC follows the same censoring scheme as in for the primary definition of PFS (Section 4.1.2). Therefore, DDC is essentially PFS of subjects whose BOR is CR or PR or SD or NON-CR/NON-PD **based on BICR-assessed BOR**.

Based on investigator-assessed BOR, DDC is defined as the time from the date of randomization to the date of the first documented tumor progression or death due to any cause in subjects whose BOR is CR or PR or SD; thus, DDC is essentially PFS of subjects whose BOR is CR or PR or SD.

4.2 Primary endpoint OS

OS is the primary endpoint for this study. It is defined as the time from the date of randomization to the date of death due to any cause in all randomized subjects. Subjects who are alive will be censored at the last known alive dates.

On treatment visits will occur at Day 1 and every 2 weeks thereafter. Subjects will be treated until unacceptable toxicity or disease progression. Following discontinuation of therapy, safety will be assessed through post-treatment Follow Up visit 2 (~ 100 days from last dose). Survival status will be assessed every 3 months after follow up visits are completed, and may be completed via telephone or in person visits. Note after discontinuation of therapy, tumor imaging assessment will continue until disease progression (see schedule of tumor imaging assessment in Section 4.1.1).

4.3 Secondary endpoints

4.3.1 ORR

The secondary endpoint of ORR is the ORR (see Section 4.1.1) in all randomized subjects with BOR determined based on BICR-assessed tumor response according to RECIST 1.1.

4.3.2 PFS

The secondary endpoint of PFS is the PFS (see Section 4.1.2) in all randomized subjects with progression is based on the first documented tumor progression assessed by BICR according to RECIST 1.1 or death.

4.3.3 PD-L1 expression

Definition of PD-L1 expression is described as follows.

PD-L1 expression missing: Subjects without an available tumor biopsy specimen for PD-L1 evaluation will be considered as PD-L1 expression missing.

For subjects with an available tumor biopsy specimen(s), the following will be considered:

PD-L1 expression is defined as the percent of tumor cell membrane staining in a minimum of 100 evaluable tumor cells per Dako PD-L1 IHC assay unless otherwise specified. This is referred as *quantifiable PD-L1 expression*. If the PD-L1 staining could not be quantified, it is further classifies as:

Indeterminate: Tumor cell membrane staining hampered for reasons attributed to the biology of the tumor biopsy specimen and not because of improper sample preparation or handling

Not evaluable: Tumor biopsy specimen was not optimally collected or prepared (e.g. PD-L1 expression is neither quantifiable nor indeterminate)

Baseline PD-L1 expression: If more than one tumor biopsy specimen is available, baseline PD-L1 expression will be determined from the most recently collected specimen (prior to randomization) with a quantifiable result. If more than one baseline PD-L1 expression measurement is available at the same day for a subject, the highest baseline PD-L1 expression measurement will be used. If all specimens for a given subject are either indeterminate or not evaluable, then the PD-L1 expression will be considered indeterminate as long as at least one specimen is indeterminate. Otherwise, PD-L1 expression will be considered not evaluable.

PD-L1 status is a dichotomized variable using an X% cut-off for quantifiable PD-L1 expression:

- PD-L1 $\geq$ X %: $\geq$ X % PD-L1 expression
- PD-L1 < X %: < X % PD-L1 expression

where X% denotes the PD-L1 expression cut-off of 1%. Additional cut off values may also be explored.

[redacted]

[redacted]

5 SAMPLE SIZE AND POWER

The sample size determination of this study is based on OS comparison between subjects randomized to receive nivolumab and sorafenib. With a total of 726 subjects randomized in a 1:1 ratio to receive nivolumab or sorafenib, approximately 91.5% power will be achieved with an overall type I error 0.05. The sample size determination is based on simulation using R v3.2 for OS. Below are details of the sample size determination.

General assumptions:

- There are assumed 39% HCV-infected subjects among all randomized subjects.
- In the sorafenib arm, median OS is 10 months for non-HCV infected and 14 months for HCV infected.
- Exponential distribution is assumed for OS in each randomized arm.

Estimated cumulated enrollment of randomized subjects is summarized in Table 5-1.

Table 5-1: Estimated Cumulative Number of Randomized Patients

Month	1	2	3	4	5	6	7	8	9	10	11	12	13
HCV	6	17	30	45	61	81	105	131	159	188	218	250	282
Non-HCV	9	25	43	67	98	132	170	210	252	297	345	394	444
All	15	42	73	112	159	213	275	341	411	485	563	644	726

Sample size determination for OS

Based on simulations under the assumptions stated above, approximately 726 subjects will be randomized and followed until at least 520 OS events are observed in order to provide 91.5% power for a hazard ratio of 0.74 with a two-sided type I error of 0.05. This accounts for a group sequential testing procedure with one interim analysis and one final analysis. The HR of 0.74 corresponds to a 35% increase in the median OS for both non-HCV infected subjects and HCV infected subjects.

The interim analysis will be conducted when 80% OS events are observed. The alpha allocation for the interim and final analyses is based on the Lan-DeMets alpha spending function approach using an O'Brien-Fleming stopping boundary controlling for a two-sided overall type 1 error of 5%. The stopping boundary will depend on the actual number of OS events observed at the time of the interim analysis and the final analysis. However, if the interim analysis is performed exactly when 80% OS events are observed, then 0.024 alpha will be used for the interim analysis and 0.043 alpha will be used for the final analysis.

It is projected that an observed hazard ratio of 0.80 or less would result in a statistically significant improvement of nivolumab at the interim analysis of OS; and an observed hazard ratio of 0.84 or less would result in a statistically significant improvement of nivolumab at the final analysis of OS.

Analysis Timing Projections

As stated above, approximately 726 subjects will be randomized to the two treatment arms in a 1:1 ratio.

It will take approximately 13 months to complete the randomization.

- OS interim analysis is projected to occur when there are at least 416 deaths (80% of target events) among approximately 726 randomized subjects, approximately 25 months after the first subject's randomization date (13 months for randomization and 12 months for survival follow-up).
- OS final analysis is projected to occur when there are at least 520 deaths among approximately 726 randomized subjects, approximately 33 months after the first subject's randomization date (13 months for randomization and 20 months for survival follow-up).

6 STUDY PERIODS, TREATMENT REGIMENS AND POPULATIONS FOR ANALYSES

6.1 Study Periods

Refer to Core Safety SAP[33] for definitions of baseline period and post baseline period for AEs and evaluations of laboratory tests, pulse oximetry and vital signs.

6.2 Treatment Regimens

The treatment group **as randomized** will be retrieved from the IVRS system:

- Nivolumab 240 mg 30 min IV
- Sorafenib 400 mg PO

The treatment group **as treated** will be the same as the treatment group randomized by IVRS. However, if a subject received the incorrect drug for **the entire period** of treatment, the subject's treatment group will be defined as the incorrect drug the subject actually received.

6.3 Populations for Analyses

All analyses will be performed using the treatment group as randomized (intent-to-treat) with the exception of extent of exposure (dosing) and safety, for which the treatment group as treated will be used. If a subject is randomized to nivolumab and receives at least 1 dose of nivolumab during the study, then the subject's treatment group is considered nivolumab. Similarly, if a subject is randomized to sorafenib and receives at least 1 dose of sorafenib, then the subject's treatment group is considered sorafenib.

When appropriate, the following definitions of different types of subjects will be applied to different populations of interest defined later in this section:

- **Enrolled Subjects:** Subjects who signed an informed consent form and were registered into the IVRS.
- **Randomized Subjects:** Enrolled subjects who were randomized to any treatment group in the study. This is the type of subjects for analyses of study conduct, study population and efficacy.
- **Treated Subjects:** Randomized subjects who received at least one dose of study drug. This is the type of subjects for analyses of exposure and safety.
- **PK Subjects:** Randomized subjects who are dosed with nivolumab and have serum time-concentration data available.
- **Response Evaluable Subjects**: Randomized subjects who have target lesion(s) assessed at baseline and at least one post-baseline timepoint.
- **PD-L1 Quantifiable Subjects**: Randomized subjects with quantifiable baseline PD-L1 expression, which exclude randomized subjects with missing or indeterminate or not evaluable baseline PD-L1 expression (see Section 4.3.3).

7 STATISTICAL ANALYSES

7.1 General Methods

Unless otherwise noted, the titles in the following subsections describe tabulations of discrete variables, by the frequency and proportion of subjects falling into each category, grouped by treatment (with total). Percentages given in these tables will be rounded and, therefore, may not always sum to 100%. Continuous variables will be summarized by treatment group (with total) using the mean, standard deviation, median, minimum and maximum values. If a missing category is not being presented in the data display, only those subjects with non-missing values for the parameter being assessed are included in the percentage calculation.

Time to event distributions (i.e. PFS, OS, DOR and DDC) will be estimated using the KM method. When appropriate, the median along with the corresponding log-log transformed 95% CI will be estimated. Rates at fixed timepoints (e.g., OS at 12 months) will be derived from the KM estimate along with their corresponding log-log transformed 95% confidence intervals.

Unless otherwise specified, the stratification factors for stratified analysis will be the following ones as entered into the IVRS:

- Etiology (HCV vs non-HCV [ie, HBV and HCC with no history of hepatitis virus infection])
- Vascular invasion and/or extrahepatic spread (present vs absent)
- Geography (Asia vs Non-Asia).

The stratified log-rank test will be performed to test the comparison of time to event distributions (OS and PFS) between two treatment arms.

The stratified hazard ratio between two treatment arms along with CI will be obtained by fitting a stratified Cox model with the treatment group variable as unique covariate.

Confidence intervals for binomial proportions will be derived using the Clopper-Pearson method[34].

The difference in rates (e.g., ORR) between the two treatment arms along with their two-sided 95% CI will be estimated using the following Cochran-Mantel-Haenszel (CMH) method of weighting[35], adjusting for the stratification factors:

$$\hat{\theta} = \frac{\sum_i w_i \hat{\theta}_i}{\sum_i w_i} \sim N\left[\theta, \frac{\sum_i w_i^2 \left[\frac{p_{ix}(1-p_{ix})}{n_{ix}-1} + \frac{p_{iy}(1-p_{iy})}{n_{iy}-1}\right]}{\left(\sum_i w_i\right)^2}\right]$$

where $\hat{\theta}_i = p_{ix} - p_{iy}$ is the rate difference of the ith stratum, $w_i = \frac{n_{ix} n_{iy}}{n_{ix} + n_{iy}}$, and n_{ix} and n_{iy} are the number of subjects randomized to treatments x and y, respectively, in the ith stratum. Associated odds-ratio will be derived.

P-values other than those provided for the primary endpoint OS and the hierarchical testing of the secondary endpoints ORR and PFS are for descriptive purpose only and not adjusted for multiplicity.

7.2 Study Conduct

7.2.1 Accrual

The accrual pattern will be summarized per country, investigational site and per month for all enrolled subjects. Randomization date, first dosing date, country, investigational site will be presented in a by-subject listing of accrual.

7.2.2 Relevant Protocol Deviations

The relevant protocol deviations will be summarized for all randomized subjects, by treatment group and overall. The following programmable deviations from inclusion and exclusion criteria will be considered as relevant protocol deviations. Non-programmable relevant eligibility and on-treatment protocol deviations, as well as significant (both programmable and non-programmable) eligibility and on-treatment protocol deviations will be reported through ClinSIGHT listings.

At Entrance:

- Subjects with ECOG PS > 1.
- Subjects with Child-Pugh B or higher.
- Subjects without evaluable disease at baseline.
- Subjects who received prior systemic therapy.
- Subjects with serum albumin < 2.8 g/dL.
- Subjects with total bilirubin > 3 mg/dL.
- Subjects with AST > 5 times the institutional upper limits of normal.
- Subjects with ALT > 5 times the institutional upper limits of normal.
- Subjects with co-infection of HBV and HCV.
- Subjects with co-infection of HBV and HDV.

On-study:

- Subjects receiving concurrent anti-cancer therapy (defined as chemotherapy, systemic therapy, radiation therapy, non-systemic therapy for HCC (local only), and surgery for HCC).
- Subjects treated differently as randomized (subjects who received the wrong treatment, excluding the never treated).

A by-subject listing will also be produced.

7.3 Study Population

Unless otherwise specified, efficacy analyses will be performed on randomized subjects with treatment group as randomized; safety analyses will be performed on treated subjects with treatment group as treated.

7.3.1 Subject Disposition

The total number of subjects enrolled (randomized or not randomized) will be presented along with the reason of not being randomized.

Number of subjects randomized but not treated along with the reason will be tabulated by treatment group as randomized. This analysis will be performed only on randomized subjects only.

Number of subjects who discontinued treatment along with corresponding reason will be tabulated by treatment group as treated.

A subject listing for all enrolled subjects will be provided showing the subject's consent date, reason for not being randomized, and reason for not being treated.

A subject listing for all treated subjects will be provided showing the subject's randomization date, first and last dosing date, off treatment date and reason for going off-treatment.

7.3.2 Demographics and Other Baseline Characteristics

The following baseline characteristics will be summarized by treatment group as randomized. All baseline presentations identify subjects with missing measurements. Listings will also be provided.

7.3.2.1 Demographics

- Age (descriptive statistics)
- Age categories (<65, ⩾65 and <75, ⩾75)
- Gender (male, female)
- Race (White, Black/African American, American Indian/Alaska Native, Native Hawaiian/Other Pacific Islander, Asian Indian, Chinese, Japanese, Asian Other, Other)
- Weight (descriptive statistics)
- ECOG Performance Status (0, 1, > 1)
- Region (US/Canada/Europe, Asia, Asia excluding Japan, Japan)

7.3.2.2 HCC Etiology

HCC etiology is a description of disease subtype at baseline and includes HBV-infected, HCV-infected, uninfected, and HBV-HCV coinfection. HCC etiology is derived based on risk factors as collected on CRF, and described below:

- Etiology is HBV-infected if HBV is a risk factor, and HCV is not included as a risk factor
- Etiology is HCV-infected if HCV is a risk factor, and HBV is not included as a risk factor
- Etiology is HBV-HCV coinfection if both HBV and HCV are included as risk factors
- Etiology is uninfected if neither HBV nor HCV is considered as a risk factor

Upon review of data, there are a few subjects with both HBV and HCV boxes checked in the risk factor CRF page. Clinical review suggested that these subjects might be similar to HCV subjects. Therefore, these subjects are grouped together with HCV subjects for all outputs of the clinical study report.

7.3.2.3 HCC Disease Characteristics

The following disease characteristics will be summarized. All available data will be listed.

- HCC Risk Factors
- Vascular invasion (Yes/No)
- Extrahepatic spread (Yes/No)
- Disease etiology (Uninfected/HCV-infected/HBV-infected/HBV-HCV coinfection)
- Child-Pugh Score (Class and Score)
- Liver Nodules (Yes/No)
- BCLC (Barcelona clinic liver cancer) stage
- AFP (descriptive statistics)
- AFP Category (≥ 200, < 200, ≥400, < 400)
- Time from initial diagnosis of HCC to randomization

7.3.3 Medical History and Pre-treatment Events

General medical history will be listed by subject. HCC or prior treatment-related pre-treatment events will be summarized by worst CTC grade, system organ class and preferred term.

7.3.4 Prior Anti-Cancer Therapies

Prior anti-cancer therapies include the following therapies and will be summarized.

- Prior systemic cancer therapy (Yes/No)
- Prior surgery therapy (Yes/No)
- Prior radiotherapy (Yes/No)
- Prior non-systemic treatment related to HCC local only (Yes/No)

7.3.5 Baseline Examinations

Subjects with abnormal baseline physical examination will be tabulated by examination criteria and by treatment group.

7.3.6 Discrepancies Between IVRS and CRF Stratification Factors

A summary table (cross-tabulation) by treatment group as randomized for stratification factor will be provided to show any discrepancies between what was reported through IVRS and CRF data.

7.4 Extent of Exposure

Analyses in this section will be performed in all treated subjects by treatment group as treated (as defined in Section 6.3). Listings will include all available exposure data.

Table 7.4-1 summarizes the key parameters used to calculate dosing data.

Table 7.4-1: Administration of study therapy: definition of parameters

	Nivolumab	Sorafenib
Dosing schedule per protocol	240mg every 2 weeks	400mg BID
Dose	*Dose (mg)* is defined as Total Dose administered. Dose administered in mg at each dosing date are collected on the CRF.	*Dose (mg)* is defined as Total Dose administered. Dose administered in mg at each dosing date are collected on the CRF.
Cumulative Dose	*Cum dose (mg) is* sum of the doses (mg) administered to a subject during the treatment period.	*Cum dose (mg) is* sum of the doses (mg) administered to a subject during the treatment period.
Relative dose intensity (%)	Cum dose (mg) / [(Last dose date - Start dose date + 14) x 240 / 14] x 100	Cum dose (mg) / [(Last dose date - Start dose date + 1) x 800] x 100
Duration of treatment	Last dose date - Start dose date +1	Last dose date - Start dose date +1

7.4.1 Administration of Study Therapy

The following parameters will be summarized by treatment group (descriptive statistics):

- Relative dose intensity (%) using the following categories: < 50%; 50 - < 70%; 70 - < 90%; 90 - < 110%; ≥ 110%.
- Number of doses received
- Cumulative dose
- Duration of treatment: duration of treatment will be presented using a KM curve whereby the last dose date will be the event date for subjects who discontinued study therapy and last dose date will be the censoring date for subjects who are still on study therapy. Median duration of treatment and associated 95% CI will be provided.
- A by-subject listing of dosing of study medication (record of study medication and dose change) and a listing of batch number will be also provided. Infusion details will also be provided for nivolumab.
- Time from randomization to first dose of nivolumab or sorafenib (0 to 3 days, 4-5 days, 6-7 days, 8-14 days, 15-21 days, 22-28 days, > 28 days).

7.4.2 Modification of Study Therapy

7.4.2.1 Dose delay for nivolumab

A dose given more than 3 days after the intended dose date will be considered a delay. Treatment may be delayed for up to a maximum of 6 weeks (42 days) from the last dose.

Length of delay is defined as duration of interval from prior dose in days minus 14 days. Dose delays will be divided into following categories: 4 - 7 days, 8 - 14 days, 15 - 42 days, 43 - 84

days, and ≥ 85 days. Dose delay events and reason for dose delay will be retrieved from CRF dosing pages.

The following parameters will be summarized for nivolumab treatment group:

- Number of subjects with at least one dose delayed, number of dose delayed per subject, length of delay, and reason for dose delay

7.4.2.2 Dose delay for sorafenib

Dose delay will occur if the subject did not receive any dose during at least one day for a different reason than "Dosing Error", provided that treatment was resumed afterwards.

Length of delay is defined as duration of interval from prior dose in days. Dose delays will be divided into following categories: <= 1 day, 2 - 3 days, 4 - 7 days, 8 - 14 days, 15 - 42 days, 43 - 84 days, and ≥ 85 days. Dose delay events and reason for dose delay will be retrieved from CRF dosing pages.

The following parameters will be summarized for sorafenib treatment group:

- Number of subjects with at least one dose delayed, number of dose delayed per subject, length of delay, and reason for dose delay

7.4.2.3 Dose modifications for nivolumab

No dose modifications for nivolumab are allowed.

7.4.2.4 Dose modifications for sorafenib

When dose reduction is necessary during the treatment of HCC, the sorafenib dose should initially be reduced to two tablets of 200 mg sorafenib once daily. For further dose reductions, refer to the sorafenib SmPC or locally approved product label for additional details.

The following parameters will be summarized for sorafenib treatment group:

- Number of subjects receiving sorafenib with at least one dose reduction and reasons for dose reduction.

7.4.2.5 Infusion interruptions and rate changes for nivolumab

The following parameters will be summarized for nivolumab:

- Number of subjects with at least one dose infusion interruption, number of infusion interruptions per subject, and the reason for interruption.
- Number of subjects with at least one IV infusion rate reduction, number of IV infusion rate reduction per subject, and the reason for reduction.

7.4.3 Concomitant Medications

Concomitant medications, defined as medications other than study medications which are taken at any time on-treatment (i.e. on or after the first day of study therapy and within 100 days following the last dose of study therapy), will be coded using the WHO Drug Dictionary.

The following summary tables will be provided:

- Concomitant medications (subjects with any concomitant medication, subjects by medication class and generic term).

A by-subject listing will accompany the table.

7.5 Efficacy

Unless otherwise specified, all analyses are based on treatment group as randomized.

7.5.1 Primary Efficacy Endpoint OS Analyses

7.5.1.1 Primary Analysis

The distribution of OS will be compared in two randomized arms via a two-sided α stratified log-rank test where α will be determined as per Section 7.5.4.2.

7.5.1.2 Secondary Analyses

A stratified Cox proportional hazards regression model will be used to estimate hazard ratio (see Section 7.1) between treatment groups along with the 100(1-α)% CI where α will be determined as per Section 7.5.4.2.

For descriptive purpose, the 95% CI for the estimated hazard ratio in Section 7.5.1.1 will be provided.

The OS curves for each treatment group will be estimated using the KM product-limit method. Median OS and the corresponding two-sided 95% confidence intervals using the log-log transformation will be computed.

Survival rates at 12, 18, 24, 36, 48 months and 5 years will be estimated using KM estimates on the OS curve for each randomized arm. Minimum follow-up (see its definition in Section 7.5.1.5) must be longer than a time period to generate a rate. Associated two-sided 95% CIs will be calculated.

The status of subjects who are censored in the OS KM analysis will be tabulated for each treatment group using following categories:

- On-study: (on treatment and not progressed, on-treatment progressed, in follow-up)
- Off-study: (lost to follow-up, withdraw consent, other)

7.5.1.3 Sensitivity Analyses

OS will be compared between treatment groups using a two-sided α (the same as in the primary analysis) unstratified log-rank test.

OS will also be compared between treatment groups using the strata as determined at baseline (CRF source) if the stratification variables at IVRS and at baseline disagree for at least 10% of the randomized subjects.

7.5.1.4 Consistency of Treatment Effect on OS in Subsets

To assess consistency of treatment effects in different subsets, a "forest" plot of the OS hazard ratio and 95% CI will be produced for the following variables, but not limited to:

- Geography (Asia vs Non-Asia)
- Region (US/Canada/Europe, Asia, Asia excluding Japan, Japan)
- Age categorization I (< 65, ≥ 65)
- Age categorization II (< 65, ≥ 65 - < 75, ≥ 75)
- Age categorization III (< 65, ≥ 65 - < 75, ≥ 75 - < 85, ≥ 85)
- Age categorization IV (< 30, ≥ 30 - < 45, ≥ 45 - < 60, ≥ 60)
- Gender (Male vs. Female)
- Race (White, Black/African American, American Indian/Alaska Native, Native Hawaiian/Other Pacific Islander, Asian Indian, Chinese, Japanese, Asian Other, Other)
- ECOG (0, 1, > 1)
- Child-Pugh Score (5, 6, (7, 8, 9 or above))
- Child-Pugh Class (A, B, C)
- Etiology (HCV-infected, HBV-infected, uninfected, HBV-HCV coinfection) from CRF
- Vascular invasion and/or extrahepatic spread (present or absent) from CRF
- AFP categorization I at baseline (AFP > 200 ng/ml, and AFP ≥ 200 ng/ml)
- AFP categorization II at baseline (AFP < 400 ng/ml and AFP ≥ 400 ng/ml)
- PD-L1 status at baseline (see Section 4.3.3)
- BCLC category at baseline

If a subgroup category has less than 10 subjects in a treatment group, then HR will not be reported for that subgroup.

7.5.1.5 Subject Follow-up

The minimum follow-up will be reported. The minimum follow-up is defined as the time interval between the last patient's randomization date and the clinical cutoff date.

The extent of follow-up defined as the time between randomization date and last known date alive (for subjects who are alive) or death date (for subjects who died) will be summarized descriptively (median, min, max) for all randomized subjects.

The currentness of follow-up for survival, defined as the time between last OS contact (ie, last known date alive or death date) and data cutoff date (defined by last patient last visit date), will be summarized by treatment group. Subjects who died before data cutoff date or subjects whose last known date alive is on or after data cut-off date will have currentness of follow-up set to zero. The currentness of follow-up will be categorized into the following categories: 0 day, 1 day-3 months, 3-6 months, 6-9 months, 9-12 months and ≥ 12 months.

7.5.1.6 Subsequent Therapy

Subsequent therapy will be summarized by treatment group and listed. The following information pertaining to subsequent therapies will be summarized:

- Systemic anti-cancer therapy by drug name
- Surgery
- Radiotherapy
- Non-systemic treatment for HCC (local only)

A by-subject listing of follow-up therapy will be produced for subjects who had any subsequent therapy.

7.5.2 Secondary Efficacy Endpoint ORR Analyses

Hierarchical testing of ORR will be performed upon demonstration of superiority in OS at OS interim or final analyses for all randomized subjects (see Section 7.5.4.2). P-values will not be presented for ORR if the endpoint OS whose null hypothesis could not be rejected. In this case, descriptive analysis and treatment difference estimations for ORR will still be presented.

Primary Analysis

ORR based on BICR assessment per RECIST 1.1 will be compared between two treatment groups using a two-sided Cochran-Mantel-Haenszel (CMH) test, stratified by the stratification factors. This comparison will only be tested if the OS comparison is positive.

Secondary Analyses

For descriptive purpose, the following statistics will be provided.

- An estimate of the differences in ORRs and corresponding 95% CI will be calculated.
- ORR along with its 95% exact CI using the Clopper-Pearson method will be computed for each treatment group.
- In addition, the stratified (source: IVRS) odds ratios (Mantel-Haenszel estimator) between the treatments will be provided along with the 95% CI.
- BOR will be summarized by response category.

Sensitivity Analyses

Similar analysis of ORR as in the secondary analyses will be performed for response evaluable subjects.

ORR might also be analyzed in some key subgroups based on patients' baseline characteristics.

7.5.3 Secondary Efficacy Endpoint PFS Analyses

PFS as secondary efficacy endpoint is defined in Section 4.3.2. Hierarchical testing of PFS will be performed upon demonstration of superiority in OS and superiority in ORR at OS interim or final analyses for all randomized subjects (see Section 7.5.4.2). P-values will not be presented for PFS if the primary endpoint OS's null hypothesis could not be rejected or the secondary endpoint ORR's null hypothesis could not be rejected. In this case, descriptive analysis and treatment difference estimations for PFS will still be presented.

Primary Analysis

PFS will be compared between the two treatment groups via the stratified log-rank test. This comparison will only be tested if the OS comparison is positive and the ORR comparison is positive.

Secondary Analyses

A stratified Cox proportional hazards regression model will be used to estimate hazard ratio (see Section 7.1) between treatment groups along with the 100(1-α)% CI (adjusted for multiplicity, see Section 7.5.4.2).

For descriptive purpose, the 95% CI for the estimated hazard ratio will be provided.

The PFS curves for each treatment group will be estimated using the KM product-limit method. Median PFS and the corresponding two-sided 95% confidence intervals using the log-log transformation will be computed.

In each treatment group, the 6, 12, 18, 24, 36, 48 months and 5 year PFS rate along with 95% CI will be estimated from the Kaplan-Meier estimate. Minimum follow-up (see its definition in Section 7.5.1.5) must be longer than a time period to generate a rate. The associated two-sided 95% CI will also be calculated.

The source of progression (death vs progression) will be summarized by treatment group.

The status of subjects who are censored in the PFS Kaplan-Meier analysis will be tabulated for each treatment group using following categories:

- Never treated
- On-study (on treatment, in follow-up)
- Off-study (lost to follow-up, withdraw consent, other).
- Received subsequent anti-cancer therapy

Sensitivity Analyses

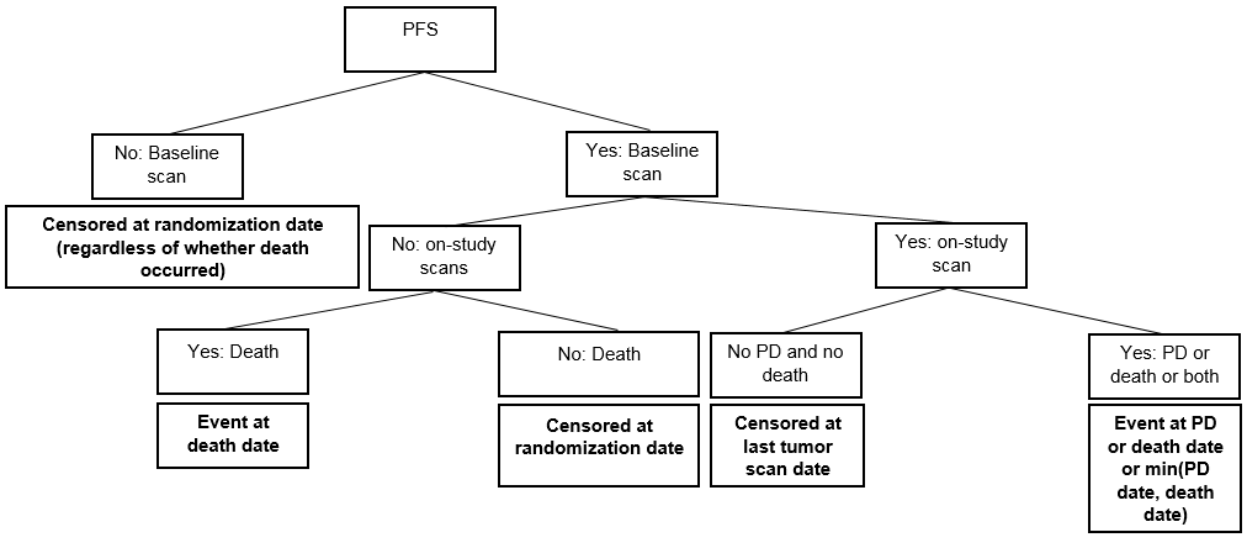
Sensitivity analyses of PFS will also be performed using the following modification of PFS primary definition.

- *PFS accounting for assessment on/after subsequent therapy* It will be defined similarly to the primary definition except that events (progression or death) and tumor assessments that occurred on or after subsequent anticancer therapy will be taken into account (see censoring scheme in Table 7.5.4-1 and Figure 7.5.4-1).

Table 7.5.4-1: Censoring Scheme - PFS Sensitivity Analysis

Situation	Date of Progression or Censoring	Outcome
No baseline tumor assessment	Date of randomization	Censored
No on study tumor assessments and no death	Date of randomization	Censored
Documented progression per BICR assessment according to RECIST 1.1	Date of the first documented tumor progression	Progressed
No progression and no death	Date of last tumor assessment	Censored
Subsequent anti-cancer therapies started without a prior reported progression	*Date of subsequent anti-cancer therapy not considered*	*None*
Death without progression and without initiation of subsequent anti-cancer therapy	Date of death	Progressed
Death without progression after initiation of subsequent anti-cancer therapy	*Date of death*	*Progressed*

Figure 7.5.4-1: Graphical Display of PFS Definition for Sensitivity Analysis



The secondary endpoint of PFS might also be analyzed in some key subgroups based on patients' baseline characteristics.

7.5.4 Interim Analyses and Multiple Comparison

7.5.4.1 Interim Analyses

An independent statistician external to BMS will perform the interim analyses, and DMC will have access to periodic unblinded interim reports of efficacy and safety.

A formal interim analysis for superiority of OS in subjects randomized to receive nivolumab vs. subjects randomized to receive sorafenib will be performed on all randomized subjects when at least 416 deaths have been observed (approximately 80% (416/520) of the total number of deaths required for the final analysis).

In addition to the formal planned interim analysis for OS, the DMC will conduct every 6 months review of safety. At each safety review, the DMC may have access to interim reports of safety and limited efficacy (such as KM curves of OS) to allow appropriate risk/benefit assessment. Details are included in the DMC charter.

The DMC will review the safety and efficacy data from the interim analysis and will determine if the study should continue with or without changes or if accrual should be stopped. Subject enrollment will continue while waiting for the DMC's decisions. More details of the interim analyses are discussed in the Data Monitoring Committee Charter.

7.5.4.2 Stagewise Hierarchical Testing Procedure for Primary Endpoint OS and Secondary Endpoints ORR and PFS

Let i = 1, 2, 3 denote index indicating primary endpoint OS and secondary endpoints ORR and PFS, respectively. Accordingly, H_1, H_2, H_3 denote two-sided null hypotheses of no effect in primary endpoint OS and secondary endpoints ORR and PFS, respectively. Let t = 1, 2 denote index indicating interim and final analysis timings, respectively. Correspondingly, $p_{i,t}$ denote nominal p-value of testing H_i at the analysis time t, and the corresponding decision boundaries can be expressed as nominal significance levels $\alpha^*_{i,t}$ such that H_i is rejected if $p_{i,t}$ is smaller than $\alpha^*_{i,t}$ at time t.

To strongly control FWER at level 0.05 across three endpoints and repeated analyses at interim and final analyses, we adopted the stagewise hierarchical testing procedure[36] to test H_1, H_2, H_3. Specifically, at the interim analysis time H_1 is tested with nominal significance level $\alpha^*_{1,1}$, leading to the following two possibilities of rejecting H_1:

- If $p_{1,1} < \alpha^*_{1,1}$, H_1 will be rejected, the trial will stop and H_2 will be tested with nominal significance level $\alpha^*_{2,1}$. If $p_{2,1} < \alpha^*_{2,1}$, H_2 will be rejected and H_3 will be tested with nominal significance level $\alpha^*_{3,1}$ with $p_{3,1} < \alpha^*_{3,1}$ leading to the rejection of H_3. If $p_{2,1} \geq \alpha^*_{2,1}$, H_2 will not be rejected and H_3 will not be tested.

- If $p_{1,1} \geq \alpha_{1,1}^*$, the trial continues to final analysis time and H_1 will be tested with nominal significance level $\alpha_{1,2}^*$. If $p_{1,2} < \alpha_{1,2}^*$, H_1 will be rejected and H_2 will be tested with nominal significance level $\alpha_{2,2}^*$. If $p_{2,2} < \alpha_{2,2}^*$, H_2 will be rejected and H_3 will be tested with nominal significance level $\alpha_{3,2}^*$ with $p_{3,2} < \alpha_{3,2}^*$ leading to the rejection of H_3. If $p_{2,2} \geq \alpha_{2,2}^*$, H_2 will not be rejected and H_3 will not be tested.

The nominal significance levels for OS $\alpha_{1,1}^*$ and $\alpha_{1,2}^*$ will be computed using EAST software based on a generalization of the Lan-DeMets error spending function approach using an O'Brien-Fleming stopping boundary given overall two-sided alpha = 0.05. The stopping boundary will depend on the actual number of OS events observed at the time of the interim analysis and the final analysis. However, if the interim analysis is performed exactly when 80% OS events are observed, then $\alpha_{1,1}^* = 0.024$ will be used for the interim analysis and $\alpha_{1,2}^* = 0.043$ will be used for the final analysis.

The nominal significance levels for ORR will be pre-specified as $\alpha_{2,1}^*$ = 0.025 and $\alpha_{2,2}^*$ = 0.025, as an equal Bonferroni splitting of the two-sided alpha = 0.05.

The nominal significance levels for PFS $\alpha_{3,1}^*$ and $\alpha_{3,2}^*$ will be computed using EAST software based on a generalization of the Lan-DeMets error spending function approach using an Pocock stopping boundary given overall two-sided alpha = 0.05. The stopping boundary will depend on the actual number of PFS events observed at the time of the interim analysis and the final analysis. 722 PFS events are expected at the final analysis time (of OS). 704 out of 722 or 97.5% PFS events are expected at the interim analysis time. If the interim analysis is performed exactly when 97.5% PFS events are observed, then $\alpha_{3,1}^* =$ 0.049 and $\alpha_{3,2}^* = 0.031$ will be used.

[redacted]

[redacted]

7.6 Safety

7.6.1 Deaths

See Core Safety SAP.[33]

7.6.2 Serious Adverse Events

See Core Safety SAP. [33]

7.6.3 Adverse Events Leading to Discontinuation of Study Therapy

See Core Safety SAP. [33]

7.6.4 Adverse Events Leading to Dose Modification

See Core Safety SAP. [33]

7.6.5 Adverse Events

See Core Safety SAP. [33]

7.6.6 Select Adverse Events

See Core Safety SAP. [33]

7.6.7 Immune modulating medication

See Core Safety SAP. [33]

7.6.8 Multiple Events

See Core Safety SAP. [33]

7.6.9 Other Events of Special Interest

See Core Safety SAP. [33]

7.6.10 Immune-Mediated Adverse Events

See Core Safety SAP. [33]

7.6.11 Laboratory Parameters

See Core Safety SAP. [33]

7.6.12 Vital Signs

See Core Safety SAP. [33]

7.6.13 Immunogenicity Analysis

See Core Safety SAP. [33]

7.6.14 Pregnancy

See Core Safety SAP. [33]

7.6.15 Adverse Events By Subgroup

See Core Safety SAP. [33] Region subgroups are formed by US/Canada/Europe vs Asia vs Asia excluding Japan vs. Japan.

7.7 Pharmacokinetics

The nivolumab concentration data obtained in this study may be combined with data from other studies in the clinical development program to develop a population PK model. This model may be used to evaluate the effects of intrinsic and extrinsic covariates on the PK of nivolumab. In addition, exposure-response analyses with selected efficacy and safety endpoints may be conducted. Results of population PK and exposure response-analyses will be reported separately.

[redacted]

7.8.1 PD-L1 Expression

The definition of PD-L1 expression and PD-L1 status is described in Section 4.3.3. Analyses of PD-L1 expression are descriptive in nature and intended to examine the distribution of PD-L1 expression and assess potential association between PD-L1 expression and efficacy measures. All the analyses will be based on baseline PD-L1 expression.

7.8.1.1 Analyses of PD-L1

Analyses of PD-L1 will include:

1. Examine the distribution of PD-L1 expression
2. Assess potential association between PD-L1 expression and efficacy measures
3. Evaluate the potential predictive relationship of the PD-L1 status and efficacy measures

1. Descriptive statistics of PD-L1 expression and PD-L1 status

- Listing of all PD-L1 IHC data, all randomized subjects
- Summary of tumor specimen acquisition and characteristics, all randomized subjects
- Summary statistics of PD-L1 expression by treatment group as randomized and overall, all PD-L1 quantifiable subjects
- Frequency of PD-L1 categorization or status (≥ 1%, < 1%), including 'missing', 'indeterminate' and 'not evaluable', by treatment group as randomized and overall, all randomized subjects

2. Evaluation of associations between PD-L1 status and efficacy measures

Analyses will be based on all randomized subjects if not otherwise specified. Each analysis will be performed for the subgroups listed below if not otherwise specified.

- Each PD-L1 status subgroup
- PD-L1 expression missing or indeterminate or not evaluable subgroup

Analyses for OS:

For each of the subgroups:

- OS curves will be estimated using the KM product limit method for each treatment group. Two-sided, 95% confidence intervals for median OS will be computed by Brookmeyer and Crowley method.

Analyses for ORR per BICR RECIST 1.1:

For each of the subgroups:

- Frequency and percentage of BOR will be summarized for each treatment group as randomized.
- ORR will be computed by treatment group as randomized along with exact 95% CIs using the Clopper-Pearson method.
- Odds ratios between the treatments will be provided along with the 95% CI.

For PD-L1 quantifiable subjects, box plots of PD-L1 expression versus response status will be provided by treatment group as randomized.

Receiver Operating Characteristics (ROC) analysis with BOR (CR+PR vs others) will be performed to help assess whether there is a clinically meaningful threshold of PD-L1 expression. For PD-L1 quantifiable subjects, plots of ROC curve by treatment group as randomized will be provided.

Analyses for PFS per BICR RECIST 1.1:

For each of the subgroups:

- PFS curves will be estimated using the KM product limit method for each treatment group. Two-sided, 95% confidence intervals for median PFS will be computed by Brookmeyer and Crowley method.

3. Evaluation of the potential predictive relationship of PD-L1 status for efficacy measures

Analyses will be based on all PD-L1 quantifiable subjects.

Analyses for OS:

When appropriate, a Cox proportional hazards regression model will be fitted for OS with treatment (treatment groups as randomized), PD-L1 status, and treatment by PD-L1 status interaction. Although the study is not designed to have appropriate power to formally test the interaction of the model, an interaction test at significance level of 0.2 will warrant further exploration and the following statistics will be reported:

- Interaction p-value
- Hazard ratio of treatment vs. control and its associated 95% CI for each of the PD-L1 status subgroup
- Hazard ratio PD-L1 ≥ 1% vs. < 1% and its associated 95% CI within each treatment group

Analyses for ORR per BICR RECIST 1.1:

When appropriate, a logistic regression model will be fitted for response status (yes=CR or PR, No=others) with treatment (treatment groups as randomized), PD-L1 status and the treatment by PD-L1 status interaction. Although the study is not designed to have appropriate power to formally test the interaction of the model, an interaction test at significance level of 0.2 will warrant further exploration and the following statistics will be reported:

- Interaction p-value
- Odds ratio of treatment vs. control and its associated 95% CI will be reported for each of the PD-L1 status subgroup
- Odds ratio of PD-L1 ≥ 1% vs. < 1% and its associated 95% CI will be reported for each treatment group

Analyses for PFS per BICR RECIST 1.1:

When appropriate, a Cox proportional hazards regression model will be fitted for PFS with treatment, PD-L1 status, and treatment by PD-L1 status interaction. Although the study is not designed to have appropriate power to formally test the interaction of the model, an interaction test at significance level of 0.2 will warrant further exploration and the following statistics will be reported:

- Interaction p-value
- Hazard ratio of treatment vs. control and its associated 95% CI for each of the PD-L1 status subgroup
- Hazard ratio PD-L1 $\geq$ 1% vs. < 1% and its associated 95% CI within each treatment group

7.9 Outcome Research Analyses

The analysis of EQ-5D and FACT-Hep data will be performed in all randomized subjects who have an assessment at baseline (Day 1, assessment prior to administration of drug on day of first dose) and at least 1 subsequent assessment while on treatment. Questionnaire completion rate, defined as the proportion of questionnaires actually received out of the expected number, will be calculated and summarized at each assessment point.

7.9.1 EQ-5D

EQ-5D data will be described by treatment group as randomized in the following ways:

- EQ-VAS scores will be summarized at each assessment time point using descriptive statistics (ie, N, mean with standard deviation and 95% CI, median, first and third quartiles, minimum, maximum).
- The proportion (N) of subjects reporting no, moderate, or extreme problems will be presented for each of the 5 EQ-5D dimensions at each assessment time point. Subjects with missing data will be excluded from the analysis.
- A by-subject listing of the level of problems in each dimension, corresponding EQ-5D health state (ie, 5-digit vector), and EQ-VAS score will be provided.

7.9.2 FACT-Hep

FACT-Hep data will be summarized by treatment group as randomized in the following way:

- Distributions of responses by category will be examined for each treatment group for the single item GP5 (“I am bothered by side effects of treatment.”). The number and proportion of subjects will be summarized for each response category at each assessment time point.
- Descriptive statistics (ie, N, mean with standard deviation and 95% CI, median, first and third quartiles, minimum, maximum) will be presented for the derived score for the following five dimensions at each assessment time point: physical well-being (PWB), social/family well-being (SWB), emotional well being (EWB), functional well-being (FWB), disease-specific hepatobiliary cancer subscale (HCS).

- FACT-Hep total scores will be summarized at each assessment time point using descriptive statistics (ie, N, mean with standard deviation and 95% CI, median, first and third quartiles, minimum, maximum).
- The number and proportion of subjects achieving a change from baseline ≥ 8 points in FACT-Hep total score will be summarized at each post-baseline assessment time point.

8 CONVENTIONS

8.1 Date Imputation

The following conventions may be used for imputing partial dates for analyses requiring dates:

- For missing and partial adverse event onset dates, imputation will be performed using the Adverse Event Domain Requirements Specification[37].
- Missing and partial Non-Study Medication Domain dates will be imputed using the derivation algorithm described in 4.3.3 of BMS Non-Study Medication Domain Requirements Specification[38].
- For death dates, the following conventions will be used for imputing partial dates:
  - If only the day of the month is missing, the 1st of the month will be used to replace the missing day. The imputed date will be compared to the last known date alive day and the maximum will be considered as the death date.
  - If the month or the year is missing, the death date will be imputed as the last known date alive day
  - If the date is completely missing but the reason for death is present the death date will be imputed as the last known date alive day
- For date of progression, the following conventions will be used for imputing partial dates:
  - If only the day of the month is missing, the 1st of the month will be used to replace the missing day. In cases where the date of death is present and complete, the imputed progression date will be compared to the date of death. The minimum of the imputed progression date and date of death will be considered as the date of progression.
  - If the day and month are missing or a date is completely missing, it will be considered as missing.
- For other partial/missing dates, the following conventions may be used:
  - If only the day of the month is missing, the 15th of the month will be used to replace the missing day.
  - If both the day and the month are missing, “July 1” will be used to replace the missing information.
  - If a date is completely missing, it will be considered as missing.

8.2 Duration Calculation and Conversion

The following conversion factors will be used to convert days to months or years: 1 month = 30.4375 days and 1 year = 365.25 days.

Duration (e.g. time from first diagnosis to first dosing date, duration of response, and time to response) will be calculated as follows:

Duration = (Last date - first date + 1)

8.3 Decimal Places

The number of decimal places displayed in all listings will be determined by the number of decimal places in the raw data.

Unless otherwise specified, minimum and maximum will be reported to the precision as the data collected, one more decimal place for the mean and median, and two more decimal places for the standard deviation. The adjusted geometric mean, geometric mean ratio and the lower and upper limits of confidence interval will be displayed to three decimal places.

8.4 Analysis Software

All statistical analyses will be carried out using SAS (Statistical Analysis System software, SAS Institute, North Carolina, USA) unless otherwise noted.

9 CONTENT OF REPORTS

The analyses described in this statistical analysis plan will be included a clinical study report except where otherwise noted. Additional exploratory analyses may be performed.

10 DOCUMENT HISTORY

Table 10-1: Document History

Version Number	Author(s)	Description
1.0	[redacted]	Initial version dated 12-28-2016
1.1	[redacted]	Added clarification on DOR censoring: Subjects who neither progress nor die will be censored on the date of their last evaluable tumor assessment. Subjects who started any subsequent anti-cancer therapy without a prior reported progression will be censored at the last evaluable tumor assessment prior to or on the date of initiation of the subsequent anti-cancer therapy. Essentially, the DOR will be censored at the same time they will be censored according to the same censor scheme as for the primary definition of PFS (Section 4.1.2). Changed definition of response evaluable subjects to the following: • Response Evaluable Subjects: Randomized subjects who have target lesion(s) assessed at baseline and at least one post-baseline

Table 10-1: Document History

Version Number	Author(s)	Description
		timepoint. Added clarification on the global study population enrollment date cutoff: A China extension is added to allow continued enrollment of subjects from China after the completion of the enrollment of the global study. The global study completed the enrollment on March 07, 2017. The enrolled subjects in the global study will be the analysis population of this statistical analysis plan (see population definitions below). Analyses related to enrolled subjects in China extension will be specified in another statistical analysis plan. Clarified the following population definitions: • **ORR Population:** Subjects who were randomized prior to or on September 26, 2016 (that is, the first 373 randomized subjects). This population will be used for the efficacy analysis of the primary endpoint of ORR. • **ORR Enrolled Population**: Subjects who were randomized prior to or on September 26, 2016 or subjects who were enrolled prior to or on September 26, 2016 but are never randomized. This is the enrollment population for the ORR population. • **All Randomized Population:** Subjects who were enrolled prior to or on March 07, 2017 and randomized. • **All Enrolled Population**: Subjects who were enrolled prior to or on March 07, 2017. Modified one existing listing and added one by-subject listing for study disposition as follows: A subject listing for all enrolled subjects will be provided showing the subject's consent date, reason for not being randomized, and reason for not being treated. A subject listing for all treated subjects will be provided showing the subject's randomization date, first and last dosing date, off treatment date and reason for going off-treatment. Added one variable for demographics: Region (US/Canada, Europe, Asia, Rest of world) Added one category for etiology: HBV-HCV coinfection. • Etiology is HBV-HCV coinfection if both HBV and HCV are included as risk factors Removed multiplicity-adjusted CI for the difference in ORRs based on CMH method of weighting.

Table 10-1: Document History

Version Number	Author(s)	Description
		Removed sensitivity analyses in the case of sources for stratification variables of IVRS and CRF baseline disagreeing at least 10%. Added Child-Pugh scores and classes as subset analyses
2.0	[redacted]	08-Nov-2017
2.0	[redacted]	Moved ORR comparison from primary objective to secondary objective. Changed ORR from one of primary endpoints to a secondary endpoint and moved its corresponding analysis plan to secondary endpoint analyses. Updated interim analysis and multiple comparison to reflect the endpoint change.
3.0	[redacted]	Replaced all randomized population with all randomized subjects Changed the categories of Region (US/Canada, Europe, Asia, Rest of world) to Region (US/Canada/Europe, Asia, Asia excluding Japan, Japan) Added the imputation algorithm for imputing investigator BOR in Section 4.4.3 Removed DCR and DDC based on Investigator Assessments according to RECIST 1.1 from exploratory endpoints Removed the analysis on concordance between BICR and Investigator Assessments Removed the description about China extension and definitions of all randomized population and all enrolled population to avoid confusion Clarified that a few subjects with both HBV and HCV boxes checked in the risk factor CRF page will be put into the HCV subgroup in the outputs of the clinical study report Added sensitivity analysis of OS in the case where the stratification variables at IVRS and at baseline (with CRF source) disagree for at least 10% of the randomized subjects Clarified that the association between OS and different progression patterns (for subjects with intrahepatic progression or extrahepatic progression) may be explored and reported external to the clinical study report