Statistical Analysis Plan

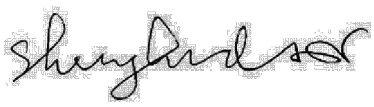
A Phase 3 Randomized, Open-Label Study Comparing Pexa-Vec (Vaccinia GM-CSF / Thymidine Kinase-Deactivated Virus) Followed by Sorafenib versus Sorafenib in Patients with Advanced Hepatocellular Carcinoma (HCC) Without Prior Systemic Therapy

STUDY PROTOCOL NUMBER:	JX594-HEP024
PHASE:	3
IND NUMBER:	BB-IND 6486
EudraCT NUMBER:	2014-001985-86
SPONSOR:	SillaJen Inc. 450 Sansome St, Suite 650 San Francisco, CA 94111- USA
COORDINATING INVESTIGATOR:	Ghassan K. Abou-Alfa, MD
DATE:	15 January, 2020

The confidential information contained in this document is the property of SillaJen, Inc. It is provided to you in confidence, for review by you, your staff, regulatory authorities, and members of ethics committees or institutional review boards. It is understood that this information will not be disclosed to any other third party, in any form, without prior authorization from the Sponsor, except to the extent necessary to obtain the written informed consent from the persons to whom the Investigational Medicinal Product may be administered.

SIGNATURES:

Author of this document
Name: Shenghua Mao
Function: Statistician
Company: PPD
City: Wilmington
Country: United States

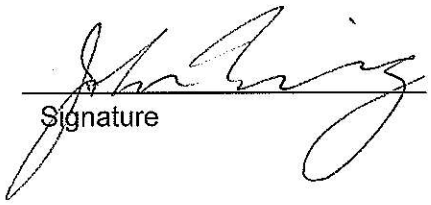


Shenghua Mao
SR BIOSTATISTICIAN
I approve this document
05 Feb 2020 23:08:30 -05:00

DocuSign

Signature Date

Sponsor representative
Name: John Ning
Function: Chief Medical Officer
Company: SillaJen Inc.
Country: United States



04-Feb-2020

Signature Date

DOCUMENT HISTORY

Version	Date	Description	Author
1.0	10-Apr-2019	Initial Release for Interim Analysis	Shenghua Mao
2.0	15-Jan-2020	Update for Final Analysis	Shenghua Mao

Table of Contents

LIST OF ABBREVATIONS

Abbreviation	Definition
AE	Adverse Event
AFP	Alfa-Fetoprotein
BOR	Best Overall Response
BSA	Body Surface Area
CI	Confidence Interval
CMD	Concomitant Medication
CR	Complete Response
CTCAE	Common Terminology Criteria for Adverse Events
DCR	Disease Control Rate
DMC	Data Monitoring Committee
DoR	Duration of Response
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic Case Report Form
EQ5D-3L	EuroQoL-5D
FACT-G	Functional Assessment of Cancer Therapy - General
FACT-Hep	Functional Assessment of Cancer Therapy - Hepatobiliary
FHSI-8	FACT Hepatobiliary Symptom Index-8
HCC	Hepatocellular Carcinoma
HR	Hazard Ratio
ICER	Incremental Cost-Effectiveness Ratio
ICF	Informed Consent Form
IMP	Investigational Medicinal Product
INT	Integer Part
IT	Intratumoral
ITT	Intent To Treat
IVRS	Interactive Voice Response System
LFU	Loss to Follow Up
MAX	Maximum
MIN	Minimum
NCI	National Cancer Institute (United States)
NE	Not Evaluable
OD	Orphan Drug
ORR	Overall Response Rate
OS	Overall Survival

Abbreviation	Definition
PD	Progressive Disease
PFS	Progression Free Survival
PP	Per Protocol
PR	Partial Response
PS	Performance Status
PSA	Probabilistic Sensitivity Analysis
PT	Preferred Term
QALY	Quality-Adjusted Life-Year
QoL	Quality of Life
(m)RECIST	(modified) Response Evaluation Criteria in Solid Tumors
SAE	Serious Adverse Event
SBP	Systolic Blood Pressure
sd	Standard Deviation
SD	Stable Disease
SI	International System of units
SLD	Sum of Longest Diameters
SOC	System Organ Class
TA	Tumor Assessment
TIR	Time to Initial Response
TSP	Time to Symptomatic Progression
TTME	Time to Tumor Markers Elevation
TTP	Time To Progression
UNK	Unknown
WHO	World Health Organization

1. UPSTREAM DOCUMENTATION

This Statistical Analysis Plan (SAP) is prepared based on the final protocol (JX594-HEP024) dated 17 February 2015, amendment 1 dated 3 October 2017, and amendment 2 dated 26 June 2019.

2. STUDY DESIGN

This is a Phase 3 multi-center, randomized, open-label study of Pexa-Vec followed by sorafenib (Arm A) versus sorafenib (Arm B) in patients with advanced hepatocellar carcinoma (HCC) without prior systemic therapy.

A total of 600 patients were to be randomized between one of two treatment arms in a 1:1 ratio (300 in each arm) to reach at least 570 evaluable patients. The final efficacy analysis was planned when 474 events (deaths) had been observed or at study early termination after the interim analysis. In addition to regular safety reviews, the interim analysis for futility was to be performed when approximately 40% (190/474) of the total planned number of deaths are documented and the details of the interim analysis for futility were provided in the separate SAP dated of 13 May, 2019. The Data Monitoring Committee (DMC) was in charge of reviewing safety data regularly during the course of the study, and safety and efficacy data at the time of interim and final analyses. The DMC reviewed unblinded data on enrollment, patient disposition, safety (e.g. mortality, adverse events, labs, vital signs, physical examination) and efficacy. Additional details are provided in the DMC Charter.

After signing the informed consent and satisfying all inclusion and exclusion criteria, patients were randomized to receive either Pexa-Vec followed by sorafenib or sorafenib alone in a 1:1 ratio (refer to Section 5.1).

In case of the discontinuation of the trial, as per the memorandum to file on interim analysis DMC meeting recommendation and study closure activities, the enrollment was supposed to be stopped at the date of discontinuation of the trial and no further treatment with Pexa-Vec to be administered. Patients randomized to Arm A and scheduled for their first treatment or awaiting re-treatment were able to initiate Sorafenib at the discretion of the principal investigator/designee provided the criteria are met. Sorafenib supply was provided until 31 Oct 2019.

In experimental Arm A, Pexa-Vec was administered as 3 bi-weekly intratumoral (IT) injections (Day 1, Week 2, Week 4). Radiological assessments were performed at screening, Week 6 and then repeated every 6 weeks until end of study treatment. At the discontinuation of study, radiographic imaging was required only for patients who transitioned onto Sorafenib prior to week 6 visit. Sorafenib was started daily at Week 6, or 2 weeks after the last IT injection, whichever is later.

In Arm B, daily sorafenib was started on Day 1, and was allowed as long as the patient

was clinically benefiting from the treatment and at least until progression or until unacceptable toxicity occurs. Radiological assessments were to be performed at Screening, on Week 6 and every 6 weeks until end of study treatment administration.

In both arms, in case study treatments are discontinued prior to the occurrence of progression, Progression Free Survival (PFS) Visits were to be performed on every 6 weeks radiological assessment.

Beyond 12 months of treatment, the radiological evaluations were to be performed every 12 weeks until first occurrence of documented progression (or study treatment discontinuation whichever is later).

When study treatments were permanently discontinued (Pexa-Vec or sorafenib), patients were to perform an End of Treatment Visit and a Safety Follow-Up Visit at least 28 days after last study treatment administration (and no more than 2 months).

For imaging data evaluations for efficacy analyses, site readers were to perform tumor assessments based on RECIST 1.1 (*Eisenhauer 2009*). In addition, independent central efficacy reads of the images were to be performed in a blinded manner with both mRECIST (modified for the purposes of the trial, as specified in the Imaging Charter) for HCC and RECIST 1.1, by a group of expert radiologists: 2 primary readers and 1 reader who acts as the adjudicator of differences between the 2 primary readers (in the case of a disagreement about the date of progression or overall response).

Up to a maximum of 5 tumors total, and a maximum of 2 tumors per organ (except for the liver where 5 tumors can be selected as non-target for this trial), representative of all involved organs will be identified as target tumors and were to be recorded and measured at baseline by the site reader. All post-baseline measurements were to be performed using the same tumors and methods as the baseline assessment.

3. STUDY OBJECTIVES

An interim analysis was conducted after 197 deaths were observed resulting in the early termination of the study due to futility based on overall survival. Therefore, the originally planned protocol efficacy analyses will be modified according to a revised regulatory strategy of Orphan Drug Designation for Pexa-Vec for the treatment of HCC, as the originally planned efficacy analyses may no longer be meaningful, while modified ad hoc analyses for efficacy endpoints to support Orphan Drug (OD) approval will be added.

3.1 OD Primary Objectives

The OD primary objective of this study is to determine radiographic responses for patients treated with Pexa-Vec followed by sorafenib (Arm A) versus Sorafenib (Arm B) based on central assessments using mRECIST for HCC for the following endpoints:

- **Time To Progression (TTP):** Time from randomization to the date of first documented radiographic tumor progression; TTP does not include deaths. If a patient has not had a TTP event at the cut-off date for analysis, TTP will be censored at the date of last evaluable tumor assessment before the cut-off.
- **Overall Response (ORR):** Proportion of patients whose best overall response during their participation in the study is either CR or PR. The best overall response is the best response recorded from the randomization until disease progression.
- **Disease Control Rate (DCR):** Proportion of patients whose best overall response during their participation in the study is either CR, PR, or SD
- **Time to Tumor Marker Elevation (TTME):** Time from randomization to the date of registered elevation of tumor marker of Alpha Fetal Protein (AFP) increase of >400 ng/ml.

3.2 OD Secondary Objectives

The OD secondary objectives of this study are:

- To determine the overall survival (OS) of 2 treatment arms.
- To determine Progression Free Survival (PFS) of 2 treatment arms.
- To determine the safety profiles of the 2 treatment arms.
- To determine the Quality of Life (QoL) of the 2 treatment arms.

3.3 Exploratory Objectives

Exploratory objectives are:

- To evaluate the effect of treatment with Pexa-Vec followed by sorafenib (Arm A) versus sorafenib (Arm B) on the following endpoints (based on central assessments for radiology endpoints):
  - Time to Initial Response (TIR): Time from randomization until the first documented response (CR or PR).
  - Tumor size over time by reference to the sum of the longest diameters of the target lesions at screening.
  - Duration of Response (DoR): Time (in months) from the first AFP decrease to disease progression (per mRECISIT/RECSIT1.1 or AFP increase>400 ng/mL) or death due to underlying cancer. To evaluate the efficacy of Pexa-Vec with respect to TTP, ORR, DCR and TTME in subgroups of patients (if the number of patients in each subgroup is sufficient) as defined in Section 8.8.3.3.
- To determine radiographic responses in the 2 treatment arms using RECIST 1.1 based on central assessments for OD primary and secondary efficacy endpoints: TTP, ORR, DCR and PFS
- To determine the Time to Symptomatic Progression (TSP) of the 2 treatment arms.
- To evaluate the efficacy of Pexa-Vec with respect to TTP and TTME in patients subdivided according to the objective response (CR or PR) at 3 months.
- To determine changes in laboratory parameters in the 2 treatment arms including standard laboratory parameters, AFP and CD4, CD8 counts.

- To evaluate the efficacy of Pexa-Vec with respect to TTP and TTME by reference to the date of introduction of sorafenib.
- To evaluate the efficacy of Pexa-Vec with respect to TTP and TTME by reference to the date of the introduction of subsequent anti-cancer treatment.

4. OVERALL STUDY DURATION

Overall study duration will consist of an active study participation phase (which includes the Treatment Phase, Long-Term Follow-Up Visits, End of Treatment Visit, Safety Follow-Up, and PFS visits (if applicable)) and a survival follow-up phase (consisting of patient and/or caregiver contact every 4 weeks).

5. TREATMENT PLAN

5.1 Randomization

Randomization was conducted in a 1:1 ratio (Pexa-Vec followed by sorafenib versus sorafenib). The randomization was performed in each subgroup of patients according to their region (Asian or Non-Asian). In each subgroup, randomization using a dynamic stochastic minimization procedure for the following factors was applied:

- Center
- Main etiology:
  - Hepatitis C,
  - Hepatitis B,
  - Alcohol,
  - Other reasons (such as hemochromatosis, Wilson's disease, type 2 diabetes, NASH)
- Presence of extrahepatic disease: Yes vs No
- Vascular invasion: Yes vs No
- Performance Status (PS): 0 vs 1
- AFP levels: <200 vs 200-400 vs >400 ng/mL

The dynamic minimization used a stochastic treatment allocation algorithm based on the variance method as proposed by Pocock and Simon (Pocock 1975).

5.2 Duration of Treatment

The duration of treatment lasts from the first study treatment administration (Pexa-Vec or sorafenib) until the last study treatment administration.

6. DEFINITION OF THE POPULATIONS TO BE ANALYZED

6.1 Intent-To-Treat (ITT) Population

The ITT population will comprise all randomized patients. Following the ITT principle, patients will be analyzed according to the treatment and stratum they were assigned to at randomization. The ITT population will be the primary population for efficacy analyses and for summaries of demographic and Baseline variables.

6.2 Safety Population

The safety population will comprise all patients who received at least one dose of study treatment (Pexa-Vec or sorafenib). The safety population will be the population for safety and drug exposure analyses and patients will be analyzed according to the treatment they received.

6.3 Per Protocol (PP) Population

The PP population will comprise all patients from the ITT population without any major protocol deviations (refer to Section 6.4) who have completed a minimum exposure requirement. In Arm A, the minimum exposure is at least one Pexa-Vec injection. In Arm B, the minimum exposure is at least 2 consecutive weeks of sorafenib. However, if a patient progressed as per Investigator radiology data, discontinued for adverse event (AE) or died before the minimum exposure requirement could be met, that patient will still be included in the PP population.

If a patient discontinued treatment for trial termination before the minimum exposure requirement could be met, that patient will still be eligible for inclusion in the PP population. Additional supply of sorafenib allowed after trial termination (i.e. 2 Aug) should not be counted for the minimum exposure, only the doses on or before August 2, 2019 will be counted in the minimum exposure for PP population.

Patients will be analyzed according to the treatment arm and stratum that they were assigned to at randomization.

6.4 Protocol Deviations

Protocol deviation will be classified as minor or major. A major deviation may lead to exclusion of patient from the PP population, only:

- If the protocol deviation is very likely to confound the scientific analysis of the primary efficacy endpoint or if it precludes any meaningful efficacy assessment,
- If it is in direct conflict with the population definition given in the title of the study (i.e., patient diagnosis, stage of disease or use of prior treatment does not correspond to the intended patient population to be studied).

The status of the major/minor protocol deviations will be reviewed by the Sponsor during the Protocol Deviations review meetings. Major protocol deviations may include, but are not limited to, the following:

- unmet inclusion exclusion criteria, at study inclusion or before each Pexa-Vec administration as described in the protocol,
- the intake of a medication contraindicated in the protocol,
- a major change in the administration schedule or dosage of Pexa-Vec or sorafenib,
- the maintenance in the study of a patient who meets study withdrawal criteria,
- a major non-compliance of a patient to the sorafenib treatment (< 80% or >120%)

7. STATISTICAL DESIGN

7.1 Sample Size Determination

Sample size was originally designed by comparing the overall survival in the Pexa-Vec arm (Arm A, Pexa-Vec followed by sorafenib) with the sorafenib arm (Arm B). Based on the aforementioned assumptions and 1:1 randomization, a total of 474 events of death should be observed to reject the null hypothesis of no Pexa-Vec effect with a power of 86% (assuming that HR = 1 for the first 6 months and 0.6 thereafter) using a stratified log-rank test at a 1-sided cumulative 2.5% level of significance.

More information about the original construction of the design with EAST® 6.3 is provided in the analysis plan prepared for the interim analysis.

As the study terminated early after first interim analysis, the final analysis will be based on actual number of subjects and events observed as of the study cut-off date for each individual subject.

7.2 Coding Dictionaries

All AEs and medical history will be coded using the most up to date version of Medical Dictionary for Regulatory Activities (MedDRA, updated twice a year). All concomitant medications will be coded using the World Health Organization Drug Dictionary (WHODRUG).

7.3 Missing Data

Measurements that were not performed or not recorded are treated as missing data. No imputation will be done for missing data, excepted for missing and incomplete dates for AE, concomitants medications (CMD), historical data, death or last contact. Missing data will be noted as missing in appropriate tables/listings.

Only birth year and age at randomization are collected. A data quality check will be programmed to ensure the duration between the randomization and birth years is equivalent to the patient's reported age, otherwise a query will be issued.

All laboratory parameters will be expressed in the International System (SI) of units. Laboratory values recorded as "value < x" or "value > x" will be handled as equal to:

- x – 0.001 if value recorded as “value < x”
- x + 0.001 if value recorded as “value > x”

For the calculation of descriptive parameters and for the value derived in standard units.

In individual listings they will be presented as reported.

Missing and incomplete dates for AE, CMD, historical data, death or last contact will be imputed as outlined below.

Imputation Rules for Partial or Missing Stop Dates for AE, CMD and medical history:

- If the month and year are present, impute the day by the last day of that month.
- If only the year is present, impute by December 31st of that year.
- If the stop date is entirely missing, assume the event or medication is ongoing.

Imputation Rules for Partial or Missing Start Dates for AE, CMD and medical history:

- If the month and year are present, impute the day by the first day of that month. If the month and year are the same as those of the date of first dose, impute the day with the day of first dose (Pexa-Vec or sorafenib).
- If only the year is present, impute by January 1st. If the year is the same as the year of the date of first dose, impute by the date of the first dose.
- If the start date is entirely missing, impute by the date of the first dose (Pexa-Vec or sorafenib).

Imputation Rules for Partial Dates for Death or Date of Last Contact:

- If the month and year are present:
  - If the month of the event is not the same as previous contact, impute the day by the first day of that month; otherwise impute the date of the previous contact.
- If only the year is present:
  - It the year of the event is not the same as previous contact, impute by January 1st of that year; otherwise impute the date of the previous contact.

The previous contact date will be defined at the last complete assessment date based on the definition of date of last contact in Section 7.4.1 using the latest complete date.

In patient data listings, dates will be displayed as reported. For the calculation of duration, the formula (end date - start date + 1) will be used.

7.4 Definitions and Derived Variables

The following conventions will be used:

- 1 month corresponds to 365.25/12 = 30.4375 days.
- 1 year corresponds to 365.25 days.

7.4.1 Definitions

Baseline

- Baseline for efficacy evaluations: the last available assessment before or at date of randomization. In the context of baseline definition, the efficacy evaluations include in particular tumor evaluations, measures for the stratification, biomarkers, and Quality of Life.
- Baseline for safety evaluations: the last available assessment before or at date of start of study treatment.

Last Contact

The last contact date will be derived for patients not known to have died at the analysis cut-off using the latest date among the following:

- All assessment dates (e.g., vital signs assessment, performance status assessment and also assessment date in third-party data such as tumor imaging, central laboratory, electrocardiogram [ECG], etc.)
- Medication dates including study medication or concomitant medications
- AE dates
- Last contact date collected
- Randomization date

The last contact date will be used for censoring of patients in the analysis of time to event.

Date of Treatment

- Start date of Pexa-Vec: the date of first injection is derived as the first date when a nonzero dose of Pexa-Vec was administered and recorded on dose administration electronic case report form (eCRF) page.
- End date of Pexa-Vec: the date of last injection is derived as the last date when a nonzero dose of Pexa-Vec was administered and recorded on dose administration eCRF page.
- Start date of sorafenib: the date of first administration of sorafenib is derived as the first date when a nonzero dose of sorafenib was administered and recorded on dose administration eCRF page.
- End date of sorafenib: the date of last administration of sorafenib is derived as the last date when a nonzero dose of sorafenib was administered and recorded on dose administration eCRF page. With the early termination of study, Sorafenib was administrated orally at the daily dose of 400mg twice daily (BID) until patient no longer clinically benefits from the treatment. SillaJen continued providing Sorafenib till Oct 31, 2019.
- Start date of study treatment: the date of first administration of study treatment is derived as the first date when a nonzero dose of any component (Pexa-Vec or sorafenib) of study treatment was administered and recorded on dosing administration eCRF page.
- End date of study treatment: the date of last administration of study treatment is derived as the last date when a nonzero dose of any component (Pexa-Vec or sorafenib) of study treatment was administered and recorded on dosing administration eCRF page. On the date of early termination of study, patients awaiting re-treatment of Pex-Vec were be given Sorafenib in the next visit (2 weeks after the last Pex-Vec). And sorafenib was continued till no longer benefit for patients or Oct 31, 2019, the last date the sponsor provided Sorafenib.

Study Day

Study Day 1 is defined as the start date of study treatment. Other study days are defined relative to the Study Day 1 with Day 2 being the day after Study Day 1 and Day -1 being the day prior to Study Day 1.

For all assessments, study day will be calculated using Day 1 as the origin, i.e., if date of assessment is on or after Day 1, the study day will be calculated as (date of assessment) – (start date of study treatment) +1; and if date of assessment is before Day 1, the study day will be calculated as (date of assessment) – (start date of study treatment).

Cut-Off Date for Analysis

Cut-off date for subjects that have not completed the study follow-up period will be based on their date of dose on or before August 2, 2019 (date of the letter sent to investigators terminating treatment + 28 days) or the start date of sorafenib administration, whichever comes earlier. The final analysis will be based on events/evaluations that occur on or before that cut-off date. The events/evaluations after that cut-off date will be included in listings, but not in analyses or summaries (eg considered as censored for time-to-event analyses). This cut-off date will also be applied for safety summaries.

Only data with an assessment date or event start date (e.g., vital sign assessment date or start date of an AE) prior to or on the cut-off date will be included in the statistical analysis. (Example: If cut-off date is 15JUN2016 then an AE starting on 15JUN2016 or before will be reported, whereas an AE with start date on 16JUN2016 or after will not be reported).

All AE or concomitant medications with start date before or on the cut-off date will be reported no matter if the end date is before or after cut-off date. If the end date is missing, then the AE will be reported as ‘ongoing’. The cut-off date will not be imputed and therefore will not appear in the listings.

If it is required to impute an end date to be able to perform a specific analysis (e.g., for a dose administration record with missing end date or end date after the cut-off date, the cut-off date will be imputed as an end date to allow for calculation of treatment exposure duration and dose intensity), the imputed date will be displayed and flagged in the listings.

Radiographic Definitions

- Overall response assessment: Overall time point tumor response is defined by either RECIST 1.1 (Eisenhauer 2009) or mRECIST for HCC. mRECIST for HCC is an adaptation of RECIST guideline which was developed for the evaluation of tumor response in patients with HCC (Forner 2009; Llovet 2008).
  - Response will be assessed with RECIST 1.1 on central evaluation.
  - Response will be assessed with mRECIST for HCC centrally only.

  More details about these evaluation methods are provided in Section 11.1.
- Disease progression: as documented on tumor assessment according to RECIST 1.1 (resp. mRECIST)
- Determination of missing adequate tumor assessments (TAs): The 'missing adequate TA' is defined as TA not done or TA with overall response equal to 'Unknown'. For the sake of simplicity, the 'missing adequate TA' will also be referred to as 'missing TA'. For PFS or TTP censoring rule, an exact rule to determine whether there is no, one or two missing TAs is needed. This rule will be based on the distance between the last adequate TA date and the event date. If the distance is larger than threshold Day 1 or Day 2 then the analysis will assume one or 2 missing TAs, respectively. The threshold Day 1 will be defined as the protocol specified interval between the TAs plus the protocol-allowed window around the assessments. Similarly, the threshold Day 2 is defined as 2 times the protocol specified interval between the TAs plus the protocol allowed window around the assessments as described in Table 1.

Table 1:
Tumor Assessment Threshold

Timing from Start Date of Study Treatment	Tumor Assessment Schedule as Per Protocol	Day 1	Day 2
Before 55 weeks	Every 6 weeks ±1 week	6+2*1 = 8 weeks	(2*6)+2*1 = 14 weeks
From 55 weeks to 67 weeks		12+2*1 = 14 weeks	(6+12)+2*1 = 20 weeks
After 67 weeks	Every 12 weeks ± 1 week	12+2*1 = 14 weeks	(2*12)+2*1 = 26 weeks

Disease Progression:

If a disease progression event is observed after 2 or more missing or non-evaluable tumor assessments, then the date of progression will be censored at the latest occurring evaluable tumor assessment; for a progression observed after a single missing or non-evaluable tumor assessment, the actual date of disease progression will be used.

Best Overall Response (BOR):

Tumor response status will be evaluated at screening and every 6 weeks until first occurrence of documented progression. Beyond 12 months of treatment, the evaluation will be performed every 12 weeks.

The best overall response is the best response recorded from the randomization until disease progression. The best overall response for each patient is determined from the sequence of overall lesion responses according to the following rules:

- Complete Response (CR): at least one determination of CR.
- Partial Response (PR): at least one determination of PR and no determination of CR.
- Stable Disease (SD): at least one determination of SD and no determination of CR or PR.
- Progressive Disease (PD): at least one determination of PD and no determination of CR, PR or SD.
- Non-Evaluable (NE): no determination of CR, PR, SD or PD.

A best overall response NE is observed in case of:

- No post-baseline assessment,
- All post-baseline assessments are NE.

Measurable disease at baseline:

Even though the eligibility criteria require measurable disease at baseline, patients without measurable disease can still be present in the data. One reason might be a simple violation of inclusion criteria. A strict adherence to the ITT principle requires including these patients in the analysis. Sensitivity analyses will be used to check the influence thereof (repeating the analysis while excluding these subjects) if more than 5% of randomized patients didn't have baseline measurable disease.

Target tumor response will always be Not Evaluated (NE) due to missing baseline

measurements. Therefore, a complete response, partial response or stable disease cannot be assigned in these cases. However, a disease progression (PD) can still be determined from non-target tumors or from new tumors.

As a result, the overall tumor responses will always be NE until PD occurs. The PFS or TTP censoring and event date options will depend on how many NEs precede the PD.

Symptomatic Progression

- Event of symptomatic progression: a decrease of 4 points or more from baseline in the Functional Assessment of Cancer Therapy (FACT) Hepatobiliary Symptom Index 8 (FHSI-8) questionnaire or a decrease in ECOG performance status to 4, or death
- Last evaluable symptomatic assessment: last assessment where either FHSI-8 or ECOG performance status are non-missing

7.4.2 Derived Variables

The following variables will be calculated. Note that the list is not exhaustive and that further variables might be added:

Duration of Treatment

- Duration of Pexa-Vec exposure (days) = (End date of Pexa-Vec treatment) – (Start date of Pexa-Vec treatment) + 14
- Duration of sorafenib exposure (days) = (End date of sorafenib treatment) – (Start date of sorafenib treatment) + 1
- Duration of study treatment exposure (days) = (End date of study treatment) – (Start date of study treatment) + 1 (if last treatment is Sorafenib) or 14 (if last treatment is Pexa-Vec).
- BSA (Mosteller formula) = Square root[Weight(kg)*Height(cm)/3600]

Definitions related to efficacy endpoints

- Time to Progression (months) = [date of first disease progression – date of randomization + 1]/30.4375 if the patient has progressed. Or [date of last evaluable

tumor assessment – date of randomization + 1]/30.4375 if the patient has not progressed (i.e., TTP time is censored).

Individual patient's TTP will be censored if no progression is observed at the cut-off date for analysis. The censoring date will be the date of the last evaluable tumor assessment before the cut-off date.

If a TTP event is observed after two or more missing or non-evaluable tumor assessments, then the date of progression will be censored at the latest occurring evaluable tumor assessment before missing assessments; for a progression observed after a single missing or non-evaluable tumor assessment, the actual date of disease progression will be used.

Patients with no post baseline tumor assessments will be censored at the randomization date for TTP.

- Time to Tumor Marker Elevation (TTME) = [date of registered elevation of tumor marker (AFP increase >400 ng/mL) – date of randomization + 1]/30.4375 if the patient has a TTME. Or [Date of last non-missing AFP value – date of randomization + 1]/30.4375 if the patient does not have TTME.
- Overall response rate (ORR) = Proportion of patients whose BOR during their participation in the study is either CR or PR.
- Disease Control Rate (DCR) = Proportion of patients whose BOR is either CR, PR or SD.
- Overall survival (months) = [Date of death from any cause – date of randomization + 1]/30.4375 if the patient has an available date of death. Or [Date of last contact – date of randomization + 1]/30.4375 if the patient does not have a date of death (i.e. survival time is censored)
- Progression free survival (months) = [MIN (date of first disease progression, date of death due to any cause) – date of randomization + 1]/30.4375 if the patient has a PFS event. Or [Date of last evaluable tumor assessment – date of randomization + 1]/30.4375 if the patient does not have PFS event (i.e., PFS time is censored)

Individual patient's PFS will be censored if no progression or death is observed at the cut-off date for analysis. The censoring date will be the date of the last evaluable tumor assessment before the cut-off date.

If a PFS event is observed after 2 or more missing or non-evaluable tumor

assessments, then the date of progression will be censored at the latest occurring evaluable tumor assessment before missing assessments; for a progression observed after a single missing or non-evaluable tumor assessment, the actual date of disease progression will be used.

Patients with no post baseline tumor assessments who do not die will be censored at the randomization date for PFS.

- Time to Initial Response (TIR): [First date of CR or PR – date of randomization + 1]/30.4375. Patients who did not achieve a response will be censored at last adequate tumor assessment date otherwise and calculated as [date of last evaluable tumor assessment – date of randomization + 1]/30.4375.
- Duration of overall response (weeks) = [MIN (date of disease progression, date of death from underlying disease) – First date of CR or PR (whichever occurs the first)) + 1]/7. Or [Date of last evaluable tumor assessment – First date of CR or PR (whichever occurs the first) + 1]/7 if the patient does not have disease progression or death from underlying disease (i.e., DoR is censored). This variable is only calculated for patients whose best overall response was CR or PR.
- Time to Symptomatic Progression (TSP): [date of first event of symptomatic progression – date of randomization + 1]/30.4375 if the patient has progressed, or [date of last evaluable symptomatic assessment – date of randomization + 1]/30.4375 if the patient does not progressed (i.e., TSP time is censored).

If a patient has not had a TSP event at the cut-off date for analysis, TSP will be censored at the date of last evaluable tumor assessment before the cut-off.

If a progression is observed after 2 or more missing or non-evaluable FHSI-8 or ECOG assessments, then the date of progression will be censored at the latest occurring FHSI-8 or ECOG assessment before missing; for a progression observed after a single missing or non-evaluable FHSI-8 or ECOG assessment, the actual date of symptomatic progression will be used.

<u>Definitions Related to Safety Endpoints</u>

Laboratory data: All laboratory data will be converted into SI units and when applicable the severity grade calculated using appropriate common terminology for adverse events (Common Terminology Criteria for Adverse Events; CTCAE version 4.03). A severity

grade of 0 will be assigned when the value is within normal limits.

- Related to study drug will include the following responses on the CRF: possibly, probably and definitely. CRF terms unrelated and remotely will be considered unrelated to study drug.

8. STATISTICAL METHODS

8.1 General Principles

Continuous variables will be presented by descriptive statistics including the number of observations (N), arithmetic mean (Mean), standard deviation (sd), minimum (MIN), median (Median) and maximum (MAX). Minimum and maximum will be presented to the same precision as the raw data. Mean and median will be presented with one more decimal place and standard deviation to 2 more decimal places than the original collected value.

Categorical variables (such as gender, race etc.) will be summarized by contingency tables containing the frequency (N) and percentage (%). Percentages will be based on all available observations.

P-values will be presented with 4 decimal places. P-values that are rounded to 0.0000 will be presented as "< 0.0001", and that are rounded to 1 will be presented as "> 0.9999".

Statistical summaries described in this plan will be produced using SAS® software version 9.2 or higher.

8.2 Patient Enrollment and Disposition

The number of screening failure patients and reasons for screening failure will be summarized. A patient listing will be provided with the reason of screening failure.

The disposition data will be presented by patient and treatment arm in data listings and the following items will be presented by treatment arm in a summary table on the ITT population:

- The number of patients randomized
- The number of patients included in each study population
- The number of patients excluded from the Per-Protocol population and reasons for exclusion
- The number of patients who have completed
  - The end of treatment visit
  - The safety follow-up visit

- The number of patients who discontinued Pexa-Vec (Arm A) and reasons for discontinuation of Pexa-Vec
- The number of patients who discontinued sorafenib (Arm A or Arm B) and reasons for discontinuation of sorafenib
- The number of patients who discontinued sorafenib before or after progression radiographic and/or clinical, whichever is first documented
- The number of patients who were followed for progression free survival after end of study treatment,
- The number of patients who were followed for overall survival after end of study treatment,
- The number of patients who discontinued the study during the follow-up for overall survival and reasons for discontinuation during this follow-up (e.g., LFU, withdrawal of consent).
- The number of subjects whose study participation was terminated due to study closure (for subjects ongoing as of August 2, 2019), which includes:
    - Patients whose study participation was terminated without transition onto Sorafenib or additional supply of Sorafenib ;
    - Patients transitioned onto Sorafenib or continued Sorafenib.

A listing of the key dates of the study will be provided including:

- The randomization dates
- The date of first, second and third Pexa-Vec injection
- The date of introduction of sorafenib
- The date of end of sorafenib
- The date of progression
- The date of death
- The date of tumor marker elevation of AFP increase > 400 ng/mL.

8.3 Analysis of Demographic and Baseline Characteristics

Baseline demographics and disease characteristics will be listed and summarized by treatment arm using the ITT population

The following quantitative data will be summarized as continuous variables:

- Age (year),
- Weight (kg),
- Height (cm),
- Body Surface Area (BSA),
- Time from randomization to initial diagnosis (months),
- Baseline tumor size (SLD) based on both mRECIST and RECIST1.1,
- AFP levels (IU/mL)

The following qualitative data will be summarized as categorical variables:

- Gender,
- Race,
- Ethnicity,
- BCLC stage,
- CLIP scoring,
- Prior therapy of HCC.

Prognostic factors used for randomization will also be summarized, including:

- Center (Asian vs Non-Asian),
- Etiology (Hep B vs Hep C vs Alcohol vs Other reasons),
- Extrahepatic disease (Yes vs No),
- Vascular invasion (Yes vs No),
- Performance status (0 vs 1) on the ECOG scale,
- AFP (<200 vs 200-400 vs >400 ng/mL).

8.4 Medical History and Current Medical Condition

Medical history will consist of any significant conditions or diseases that stopped at or prior to the time of signing the informed consent (IC). Current medical conditions are significant condition started before IC with ongoing condition, or diseases present at time of informed consent through end of study.

The number and percentage of patients with relevant medical history/current medical conditions (coded per MedDRA) will be presented by treatment arm, system organ class (SOC) and preferred term (PT) in the ITT population. A listing of patients with relevant medical history/current medical conditions will also be provided.

8.5 Minimization/Stratification Factors

Minimization factors (main etiology, presence of extrahepatic disease, vascular invasion, ECOG performance status, AFP levels) are used in randomization to minimize imbalance among groups with respect to both the number in each treatment arm and the characteristics of each treatment group. Those factors will not be used in the primary efficacy analysis and will be treated as covariates for secondary or sensitivity analyses. Region is the only statistical stratification factor and will be used in primary, secondary and sensitivity analyses. Frequencies for each stratification and minimization factor collected in the IVRS will be summarized by treatment arm.

Baseline prognostic factors collected on the eCRF will be cross-classified, tabulated and listed against the prognostic factors used to randomize patients. If more than 5% of randomized patients are misclassified according to the prognostic factors used in the minimization, there will be an additional sensitivity analysis for TTP and TTME which will utilize the Cox proportional hazards model, stratified by region with the other minimization factors, according to the actual value of the characteristic as recorded in the EDC, as covariates.

8.6 Protocol Deviation Summaries

The number and percentage of patients in the ITT population and excluded from PP population with protocol deviations will be tabulated by treatment arm. The protocol deviations will also be summarized by center.

All protocol deviations will be listed by treatment arm.

8.7 Treatments

8.7.1 Prior and Concomitant Medications

Any medications (as well as HCC medications) or significant non-drug therapies (such as prior radiation therapy, prior HCC surgery, prior HCC local regional therapy) starting and ending before Day 1 will be defined as prior medication. Any non-study medication or any therapeutic intervention (e.g., surgery, blood transfusion) with a start date is on or after Day 1 up to 28 days after the last dose of Pexa-Vec or sorafenib (e.g., up to the safety follow-up visit)., inclusive, or with a start date before Day 1 and an end date after Day 1 or ongoing, will be considered concomitant medication.

All non-study medication will be coded by Anatomical Therapeutic Chemical (ATC) code and PT using the latest version of World Health Organization Drug Dictionary (WHODrug) and be summarized in frequency tabulations (subject counts and percentages) by ATC forth level, PT, and treatment arm.

A listing of non-study medication will also be provided, to include dose, unit, frequency, route of administration, start and end dates, and reason for use.

The ITT population will be used for all summaries and listings.

8.7.2 Further Anti-Cancer Therapies

All anti-cancer therapies given after progressive disease will be coded using the latest WHODrug version. For patients who prematurely discontinued the study treatment (for another reason other than progressive disease), all further antineoplastic therapies will also be collected and coded in the same way. Anti-
cancer therapies received after the last dose of study medication are recorded at follow up visits after end of treatment (EOT), which also include concomitant treatment before t he initiation of survival follow up visits but after EOT date.

All anti-cancer therapies will be listed and summarized by active ingredient and treatment arm by means of frequency counts and percentages using the ITT population.

8.7.3 Study Treatment

Number of Pexa-Vec doses received (partial doses will be counted as one dose) will be categorized (0, 1, 2 or 3). Cumulative injection volume for all the injections, and average injection volume received for each treatment session will also be summarized. For the calculation of average injection volume, a single average value will be computed for each patient, and then the average will be computed for each treatment arm.

Average injection volume received = Cumulative injection volume for all the injections/ Number of Pexa-Vec doses received.

Exposure of sorafenib will be provided by summarizing the duration of exposure and the average daily dose (mg/day) by treatment arm. Number of patients who received the planned dose and who had a dose modification during the study will be summarized. For the calculation of daily dose (mg/day), a single average value will be computed for each patient, and then the average will be computed for each treatment arm.

Duration of exposure = End date of sorafenib - Start date of sorafenib+1.

Average daily dose = Cumulative dose taken / Duration of exposure.

In addition, total duration of study treatment (including Pexa-Vec and sorafenib) will be listed and summarized by treatment arm.

Total duration of study treatment = End date of study treatment - Start date of study treatment+1 (or +14 for Pexa-Vec subjects discontinuing prior to receiving Sorafenib).

A patient's study drug compliance (%) to sorafenib will be calculated as {(number of tablets dispensed - number of tablets returned)/ (number of tablets dispensed)}*100%. This compliance will be summarized with descriptive statistics and patient counts within compliance categories (<80%, 80 to <100%, 100 to <120%, >=120%). If the number of returned tablets and/or the return date are missing, then 100% compliance will be assigned for each day up to the number of tablets dispensed, the date of return, or the date of last dose, whichever is earliest. The safety population will be used for all summaries and listings of study treatment.

8.8 Analysis of Efficacy

Efficacy analyses will be performed on the ITT population. OD Primary and OD secondary endpoints will be repeated using the PP population. A re-randomization test will be

performed to obtain one-sided p value, and the re-randomization should be undertaken 1000 times. Section 11.3 provides details of re-randomization algorithm applied in this study.

Analyses based on radiological endpoints will be performed based on central assessments using mRECIST for HCC and will be repeated using central RECIST 1.1 criteria.

8.8.1 OD Primary Efficacy Endpoint and Analysis

Primary Analysis

The primary objective of this study is to determine radiographic responses in the 2 treatment arms based on central assessments using mRECIST for HCC for the following endpoints: Time to progression (TTP), Overall Response (ORR), Disease Control Rate (DCR) and Time to Tumor Marker Elevation (TTME).

8.8.1.1 Time To Progression (TTP)

TTP is defined as the time from randomization to the date of first documented radiographic tumor progression; TTP does not include deaths. If a patient has not had a progression event at the cut-off date for analysis, will be censored at the date of last evaluable tumor assessment before the cut-off. If a progression event is observed after 2 or more missing or non-evaluable tumor assessments, then the date of progression will be censored at the latest occurring evaluable tumor assessments; for a progression observed after a single missing or non-evaluable tumor assessment, the actual date of disease progression will be used.

A re-randomization test using stratified log rank test stratified by region will be performed to compare the 2 treatment arms. Estimates of the HRs (Arm A over Arm B) with 95% CI will be obtained from a Cox Proportional Hazard (PH) model stratified by region with other minimization factors as covariates will also be presented.

TTP will be presented descriptively for each treatment arm separately using Kaplan-Meier curves. Summary statistics from the Kaplan-Meier distributions will be determined, including median TTP and 25% and 75% quartiles with corresponding 95% confidence intervals (CIs). The proportions of patients remaining progression free at 3, 6, 9 and 12 months, along with 95% CIs will also be provided by treatment arm.

Supportive Analysis

Estimates of the HRs (Arm A over Arm B) with 95% CI for TTP will be obtained from a Cox proportional hazard (PH) model stratified by region with, etiology, extrahepatic disease, vascular invasion, ECOG performance status and AFP level as covariates in the model in the ITT population.

- In order to take the delayed effect into account, the model will allow for a time dependent treatment effect which will take the form of a binary time indicator (≤ 6 months and > 6 months). An interaction test will be conducted to assess the statistical significance of the time dependent effect.

Sensitivity Analyses

- An unstratified re-randomization test will be performed to assess the robustness of the results. This test will utilize the unstratified log-rank statistic. Hazard ratios (HR), together with associated 95% CI, resulting from an unstratified Cox model but with region and the remaining minimization factors included in the model as covariates will also be presented in the ITT population.
- All analyses for TTP will be repeated except that TTP will not be censored if a progression is observed after 2 or more missing or non-evaluable tumor assessments.

8.8.1.2 Overall Response Rate (ORR)

Overall Response Rate (ORR) is defined as the proportion of patients whose Best Overall Response (BOR) is either CR or PR, patients with NE response will be counted in the denominator.

ORR will be presented by treatment arm along with exact 95% CIs. Difference in ORR proportions (with 95% CI) will also be provided. A Cochran-Mantel-Haenszel test will be performed to compare the 2 treatment arms with respect to the ORR at a 1-sided 2.5% level of significance. A re-randomization test stratified for region based on the Mantel-Haenszel Chi-square statistic will be performed to obtain one-sided p-value.

As a sensitivity analysis, a two-sided Wilson Newcombe 95% CI for the difference in ORR, stratified by region will be computed by using the method of Yan and Su (*Yan 2010*).

Patients with BOR 'NE' will be summarized by reason for having unknown status.

8.8.1.3 Disease Control Rate (DCR)

Disease Control Rate (DCR) is defined as the proportion of patients whose BOR is either CR, PR or SD.

DCR will be presented by treatment arm along with exact 95% CIs. Difference in proportions (with 95% CI) will also be provided. A Cochran-Mantel-Haenszel test will be performed to compare the 2 treatment arms with respect to the DCR at a 1-sided 2.5% level of significance. A re-randomization test stratified for region based on the Mantel-Haenszel Chi-square statistic will be performed to obtain one-sided p-value.

As a sensitivity analysis, a two-sided Wilson Newcombe 95% CI for the difference in DCR, stratified by region will be computed by using the method of Yan and Su (*Yan 2010*).

8.8.1.4 Time to Tumor Marker Elevation (TTME)

Time to tumor marker elevation is defined as time from randomization to the date of registered elevation of tumor markers (AFP) with increase of >400 ng/mL. If a patient is alive or if the patient has no tumor marker increase of >400 ng/mL at the cut-off date for analysis, TTME will be censored at the date of last AFP record before the cut-off.

TTME will be summarized by treatment arm. A Kaplan-Meier curve will be constructed for each treatment arm. Median TTME and 25% and 75% quartiles will be presented along with 95% CIs for each treatment arm. In addition, the Kaplan-Meier estimates with 95% CIs at 3, 6, 9 and 12 months will be presented by treatment arm. Numbers of patients at risk will also be displayed at monthly intervals below the time axis for each of the two treatment groups and censored observations will be marked by notches on the curves.

Supportive Analysis

Estimates of the HRs (Arm A over Arm B) with 95% CI for TTME will be obtained from a Cox proportional hazard (PH) model stratified by region with, etiology, extrahepatic disease, vascular invasion, ECOG performance status and AFP level as covariates in the model in the ITT population.

- In order to take the delayed effect into account, the model will allow for a time dependent treatment effect which will take the form of a binary time indicator (≤ 6 months and > 6 months). An interaction test will be conducted to assess the statistical significance of the time dependent effect.

Sensitivity Analysis

- An unstratified re-randomization test will be performed for TTME to assess the robustness of the results. This test will utilize the unstratified log-rank statistic.
- Hazard ratios (HR), together with associated 95% CI, resulting from an unstratified Cox model but with region and the remaining minimization factors included in the model as covariates will also be presented in the ITT population.
- Analysis for TTME will be repeated by excluding patients with two consecutive missing AFP values before cut-off date in ITT population.

8.8.2 Secondary Efficacy Endpoints and Analyses

Secondary efficacy endpoint analyses include overall survival (OS) analysis, progression free survival analysis (PFS).

8.8.2.1 Overall Survival Analysis (OS)

Overall survival between Arm A (Pexa-Vec followed by sorafenib) and Arm B (sorafenib) will be compared using a stratified re-randomization test. This test will utilize the stratified log-rank statistic stratified by region.

Overall survival will be presented descriptively for each treatment arm using a Kaplan-Meier curve. Summary statistics from the Kaplan-Meier distribution will be determined, including median overall survival and 25% and 75% quartiles with corresponding 95% CIs. The proportions of patients alive at 6, 9, 12, 18, and 24 months, along with corresponding 95% CIs, will also be provided by treatment arm in the ITT population. Numbers of patients at risk will also be displayed at monthly intervals below the time axis for each of the two treatment groups and censored observations will be marked by notches on the curves

Handling of Missing Values/Censoring/Discontinuations

The survival status of all patients will be determined at the cut-off date. Patients without a known date of death on or before the cut-off date will be censored at the date of their last contact.

If a patient withdraws early, overall survival will not be censored at the date of withdrawal unless this is the date they were last known to be alive.

The number of overall survival observations that are censored will be summarized by treatment arm, according to the reason for censoring.

Supportive Analysis

Estimates of the HRs (Arm A over Arm B) with 95% CI will be obtained from a Cox proportional hazard (PH) model stratified by region with, etiology, extrahepatic disease, vascular invasion, ECOG performance status and AFP level as covariates in the model in the ITT population.

Sensitivity Analysis

- An unstratified re-randomization test will be performed to assess the robustness of the results. This test will utilize the unstratified log-rank statistic.
- Hazard ratios (HR), together with associated 95% CI, resulting from an unstratified Cox model but with region and the remaining minimization factors included in the model as covariates will also be presented in the ITT population.

8.8.2.2 Progression Free Survival (PFS)

PFS is defined as the time from randomization to the date of first documented radiographic tumor progression or death due to any cause, whichever occurs first. If a patient has not had a PFS event at the cut-off date for analysis, PFS will be censored at the date of last evaluable tumor assessment before the cut-off. If a PFS event is observed after 2 or more missing or non-evaluable tumor assessments, then the date of progression will be censored at the latest occurring evaluable tumor assessments; for a progression observed after a single missing or non-evaluable tumor assessment, the actual date of disease progression will be used.

A re-randomization test using stratified log rank test stratified by region will be performed

to compare the 2 treatment arms. Estimates of the HRs (Arm A over Arm B) with 95% CI will be obtained from a PH model stratified by region with other minimization factors as covariates will also be presented.

PFS will be presented descriptively for each treatment arm separately using Kaplan-Meier curves. Summary statistics from the Kaplan-Meier distributions will be determined, including median PFS and 25% and 75% quartiles with corresponding 95% CIs. The proportions of patients remaining progression free at 3, 6, 9, 12, 18 and 24 months, along with 95% CIs will also be provided by treatment arm.

Sensitivity Analyses

- An unstratified re-randomization test will be performed to assess the robustness of the results. This test will utilize the unstratified log-rank statistic. Hazard ratios (HR), together with associated 95% CI, resulting from an unstratified Cox model but with region and the remaining minimization factors included in the model as covariates will also be presented in the ITT population.
- Analyses will be repeated in the same way except that PFS will not be censored if a progression is observed after 2 or more missing or non-evaluable tumor assessments.

8.8.3 Exploratory Efficacy Endpoints and Analyses

Analyses performed based on central assessments using mRECIST for HCC will be repeated using RECIST 1.1 based on both local and central assessments in the ITT population.

8.8.3.1 Time to Initial Response

The Time to Initial Response (TIR) is defined as the time from randomization until the first documented response (CR or PR). Patients who did not achieve a response will be censored at last adequate tumor assessment date otherwise.

TIR will be summarized by treatment arm. A Kaplan-Meier curve will be constructed for each treatment arm. Median TIR and 25% and 75% quartiles will be presented along with 95% CIs for each treatment arm. In addition, the Kaplan-Meier estimates with 95%

confidence intervals at 3, 6, 9 and 12 months will be presented by treatment arm. Number of patients at risk will also be displayed at monthly intervals below the time axis for each of the two treatment groups and censored observations will be marked by notches on the curves.

8.8.3.2 Time to Symptomatic Progression (TSP)

Time to Symptomatic Progression (TSP) is defined as the time from randomization until the first documented event of symptomatic progression denoted as a decrease of 4 points or more from baseline in the Functional Assessment of Cancer Therapy (FACT) Hepatobiliary Symptom Index 8 (FHSI-8) questionnaire or a decrease in ECOG performance status to 4, or death. If a patient has not had a TSP event at the cut-off date for analysis, TSP will be censored at the date of last evaluable tumor assessment before the cut-off. If a progression is observed after 2 or more missing or non-evaluable FHSI-8 or ECOG assessments, then the date of progression will be censored at the latest occurring FHSI-8 or ECOG assessment before missing; for a progression observed after a single missing or non-evaluable FHSI-8 or ECOG assessment, the actual date of symptomatic progression will be used.

TSP will be presented descriptively for each treatment arm separately using Kaplan-Meier curves. Summary statistics from the Kaplan-Meier distributions will be determined, including median TSP and 25% and 75% quartiles with corresponding 95% CIs. The proportions of patients remaining progression free at 3, 6 and 9 months along with 95% CIs will also be provided by treatment arm.

8.8.3.3 Duration of Response (DoR)

Duration of Response (DoR) applies only to patients whose best overall response is CR or PR. The DoR is defined as the time (in months) from the first AFP decrease to disease progression (per mRECISIT/RECSIT1.1 or AFP increase>400 ng/mL) or death due to underlying cancer. If a patient is alive or if the cancer has not progressed at the cut-off date for analysis, DoR will be censored at the date of last evaluable tumor assessment before the cut-off.

If a progression is observed after 2 or more missing or non-evaluable tumor assessments, then the date of progression will be censored at the latest occurring evaluable tumor assessment before missing assessments; for a progression observed after a single

missing or non-evaluable tumor assessment, the actual date of disease progression will be used.

DoR will be summarized by treatment arm. A Kaplan-Meier curve will be constructed for each treatment arm. Median DoR and 25% and 75% quartiles will be presented along with 95% CIs for each treatment arm. In addition, the Kaplan-Meier estimates with 95% CIs at 3, 6, and 9 months will be presented by treatment arm. Numbers of patients at risk will also be displayed at monthly intervals below the time axis for each of the two treatment groups and censored observations will be marked by notches on the curves.

8.8.3.4 Tumor Size Over Time

Tumor size is defined as the sum of the longest diameters (SLD), calculated by the sum of the longest diameters (LDs) of viable enhancing hepatic target tumors plus LDs of any non-nodal extrahepatic target tumors plus the short axis diameters of any nodal target tumors. For each time point, SLD will be calculated along with the relative change from baseline.

A repeated measurements analysis model (implemented via SAS PROC MIXED), that includes terms of treatment arm, baseline stratification factors, baseline value and visit (as a factor) as fixed effect, with a general (unstructured) residual variance-covariance matrix, will be used to compare the 2 treatment arms with respect to changes in the tumor size longitudinally over time. Treatment*visit interaction will be used to investigate the possibility of time-dependent treatment effect. A heterogeneous order-1 autoregressive variance-covariance structure will be used in case of problems of convergence of the model with a general structure. Means estimated by the model (with 95% CI associated) at each visit will be summarized for each treatment arm. In case of interaction Treatment*Visit is significant, Tukey post-hoc tests, to adjust for p-values and confidence intervals for visit comparisons to baseline, will be performed to compare the 2 arms at a given visit.

Waterfall plots will be given per treatment arm. These plots will display the best percentage change from baseline in the sum of the longest diameter of target tumors for each patient.

8.8.3.5 Efficacy in Subgroups of Patients

The purpose of these subgroup efficacy analyses is to assess the consistency of

treatment effect across potential or expected prognostic factors. If there are too few events for a meaningful analysis of a particular subgroup (i.e., less than 10 events in a subgroup), the relationship between that subgroup and endpoint will not be analyzed and descriptive summaries will be provided instead.

The homogeneity of the treatment effect for Pexa-Vec will be evaluated in subgroups of patients (if the number of patients is sufficient, see above) according to the following criteria:

Baseline categorical variables:

- Region (Asia or non-Asia),
- Countries,
- Main etiology,
- Presence of extrahepatic disease,
- Performance Status (PS) on ECOG scale (0, 1),
- AFP levels (<200 or ≥200 ng/mL, <400 or ≥400 ng/mL),
- Gender,
- Macroscopic vascular invasion,
- Cirrhosis,
- BCLC stage (B or C),
- Child Pugh (A5 or A6),
- Prior therapy of HCC,
- Histological grade.

Post-Baseline categorical variables:

- Whether subject received anti-cancer therapy after last dose of study medication (based on those therapies recorded on the survival follow-up CRF) (Yes or No)
- Type of anti-cancer therapy type received after discontinuation of study treatment (top 3 types)
- Number of Pexa-Vec doses received (1, 2 or 3)
- Exposure duration of Sorafenib (quartiles)

- Treatment compliance to Sorafenib (<80%, 80 to <100%, >=100%).

Continuous baseline variables:

- Age,
- Baseline SLD,
- AFP,
- Time since initial diagnosis.

For continuous variables, subgroups will be defined using quartiles (Q1, Median or Q3) calculated using baseline values measured in all randomized patients.

In these subgroups, TTP and TTME will be analyzed using a Cox PH model (stratified by Region, except when looking at the factors for Region and for country) including only the treatment term as an independent variable. In addition, a forest plot presenting HR (with 95% CI) will be provided for each subgroup. For each of the factors listed above, a p-value for interaction will be obtained from the stratified Cox PH model (stratified by Region, except when looking at the factors for Region and for country) by including terms for treatment, factor and treatment by factor interaction. These p-values for interaction will be displayed on the Forest plot.

Subgroup analysis will be performed for ORR and DCR using Cochran-Mantel-Haenszel test (stratified by Region, except when looking at the factors for Region and for country) including only the treatment term as an independent variable to compare the 2 treatment arms at a 1-sided 2.5% level of significance. In addition, a forest plot presenting difference in rates (with 95% CI) will be provided for each subgroup.

Subgroups based on compliance and exposure will be summarized for the corresponding treatment group only.

Subgroup analyses will be performed in both ITT and PP population.

8.8.3.6 Efficacy with Other Date of Introduction as a Reference

Efficacy of Pexa-Vec will be evaluated with respect to TTP and TTME by reference to the date of introduction of sorafenib (instead of the date of randomization). Patients who did

not receive sorafenib will be excluded from the analysis.

Efficacy of Pexa-Vec respect to TTP and TTME will also be evaluated by reference to the date of introduction of subsequent anti-cancer treatment (instead of the date of randomization) in the same way as specified above. Subjects with events occurring prior to induction of subsequent anti-cancer therapy will be excluded from the analysis

A Kaplan-Meier curve will be constructed for each treatment arm. Median TTP and TTME and 25% and 75% quartiles will be presented along with 95% CIs for each treatment arm. In addition, the Kaplan-Meier estimates with 95% CIs at 3, 6, 9 and 12 months will be presented by treatment arm. Numbers of patients at risk will also be displayed at monthly intervals below the time axis for each of the two treatment groups and censored observations will be marked by notches on the curves.

Analysis will also be conducted in PP population.

8.8.3.7 Efficacy with Respect to Objective Response

Efficacy of Pexa-Vec will be evaluated with respect to TTP and TTME in patients subdivided according to the presence or not of an objective response (CR or PR at 3 months

Estimates of the HRs (Arm A over Arm B) with 95% CI will be obtained from a Cox Proportional Hazard (PH) model stratified by region and according to objective response (CR or PR) at 3 months.

Analysis will also be conducted in PP population.

8.8.3.8 Clinical and Efficacy Laboratory Evaluations

Clinical laboratory parameters include AFP and CD4/CD8 counts.

In addition, during the study, samples will be archived to further assess immune parameters and identify potential biomarkers or biomarker ‘profiles’ of patient populations most likely to benefit from treatment with Pexa-Vec.

Relationship between Pexa-Vec efficacy and clinical laboratory parameters at single time points and over time will be described.

These analyses will include but will not be restricted to:

- The investigation of the correlation between parameters at baseline
- The correlation of each baseline or post baseline value with clinical endpoint
- Descriptive changes of parameters from baseline to post treatment
- Correlation of the change from baseline with clinical endpoint

8.9 Analysis of Safety

Safety analyses will be based on the safety population. The safety summary tables will include all safety assessments collected from the first administration of any study treatment up to 28 days after end date of any study treatment. All safety data will be listed and those collected later than 28 days after end date of study treatment will be flagged in the listings.

8.9.1 Adverse Events

All AE tables will be presented with number and percentage of patients and number of events for each treatment arm. The percentage of patients will be based on all treated patients within the treatment arm. The numbers on which the percentages are based will be indicated in the headings of the columns.

Data will be presented by SOC and PT using MedDRA. MedDRA SOC and PT terms within system organ class will be sorted by descending frequency in the overall group. PTs or SOCs with the same frequency will be sorted alphabetically.

The CTCAE grades will always be displayed in the sequence: Total, Grade 1 (Mild), Grade 2 (Moderate), Grade 3 (Severe), Grade 4 (Life-threatening) and Grade 5 (Death).

A patient with multiple occurrences of an AE will be counted only once in the AE category.

A summary of treatment-emergent adverse events will be provided by treatment arm including the number (and percentage) of patients with:

- At least one AE, AE related to Pexa-Vec only, AE related to sorafenib only, AE related to both sorafenib and Pexa-Vec, AE related to Pexa-Vec (no matter if related to sorafenib) or AE related to sorafenib (no matter if related to Pexa-Vec)
- At least one SAE, SAE related to Pexa-Vec only, SAE related to sorafenib only, SAE related to both sorafenib and Pexa-Vec, SAE related to Pexa-Vec (no matter if related to sorafenib), SAE related to sorafenib (no matter if related to Pexa-Vec)
- At least one AE leading to discontinuation

All AEs, AEs leading to discontinuation of study treatment (overall and by component: Pexa-Vec or sorafenib), serious AEs (SAE) will be listed and summarized by SOC and PT, intensity (based on the NCI CTCAE v4.03), relationship to study treatment (Pexa-Vec only, sorafenib only, at least to Pexa-Vec or at least to sorafenib), relationship to IT procedure (intratumoral administration of Pexa-Vec) and treatment arm. In patients with multiple occurrences of the same AE, the AE with the maximum grade will be presented in table. An additional table will be presented for AEs leading to dose modification of sorafenib.

Written narratives will be produced for all SAEs and pregnancies and unexpected or other important AEs that are judged to be of special interest because of their clinical importance. For example, any problem during the pregnancy and/or its outcome should be reported as an AE or a SAE. Spontaneous or induced abortions as well as ectopic pregnancy should be considered as serious. Any problem concerning the newborn should also be reported as an AE or SAE.

SAEs occurring after signing the Inform Consent Form (ICF) but before starting study treatment, including those observed in patients randomized but never treated with the Investigational Medicinal Product (IMP), will be listed separately from those occurring after treatment start.

Individual AE listings will be provided including patient identifiers, all data given on the AE CRF page (including the verbatim term), MedDRA system organ class, MedDRA preferred term, duration of AE (stop date - start date + 1, or ongoing) and the study day of AE start (start date – day 1 date + 1). Listings for AEs will be sorted by treatment group, patient identifier, start date and preferred term.

8.9.2 Deaths

AEs leading to death will be listed and summarized by System Organ Class (SOC), Preferred Term (PT), and treatment arm, relationship to study treatment (Pexa-Vec or sorafenib) and relationship to procedure (IT administration of Pexa-Vec).

A summary of AEs leading to death will be provided by treatment arm including the number (and percentage) of patients with:

- At least one AE leading to death, AE leading to death related to Pexa-Vec only, AE leading to death related to sorafenib only, AE leading to death related to both sorafenib and Pexa-Vec, AE leading to death related to Pexa-Vec (no matter if related to sorafenib), AE leading to death related to sorafenib (no matter if related to Pexa-Vec)

A listing will also be provided.

8.9.3 Laboratory Abnormalities

The summaries will include all laboratory assessments collected from the first administration of any study treatment and no later than 28 days after end date of study treatment. All laboratory assessments will be listed and those collected later than 28 days after end date of study treatment will be flagged in the listings.

All laboratory values will be converted into SI units and when applicable, the severity grade calculated using appropriate common terminology criteria for AEs (CTCAE, version 4.03). A severity grade of 0 will be assigned when the value is within normal limits.

A listing of laboratory values will be provided by laboratory parameter, patient and treatment arm. A separate listing will display notable laboratory abnormalities (i.e., newly occurring CTCAE Grade 3 or Grade 4 laboratory toxicities). The frequency of laboratory abnormalities will be displayed by parameter and treatment arm using the following summaries:

- A Shift table using CTCAE grades to compare baseline to the worst post-baseline value will be produced for hematology and biochemistry laboratory parameters with CTCAE grades. Note that for parameters with 2 directions abnormalities (hypo/hyper), both will be presented.

8.9.4 Other Laboratory Data

Other laboratory data (e.g., vital signs, ECG, ECOG, AFP, CD4/CD8 count) will be listed and summarized using descriptive statistics by visit. Change from baseline will also be included in the descriptive statistics summary. In addition, shift tables from baseline to subsequent visits will be presented for ECG, ECOG.

Vital sign assessments are performed in order to characterize basic body function. The parameters collected in this study are: height (cm), weight (kg), heart rate (beats per

minute) and systolic (SBP) and diastolic blood pressure (DBP) (mmHg).

Proportion of patients who have SBP drop below 90, 80, 70, and 60 mmHg after IT session will be summarized at each visit and during the study.

Proportion of patients who have fever peak above 38°C and 39°C after IT session will be summarized at each visit and during the study. Time to fever and duration of fever above 38°C and 39°C after each IT session will be summarized using descriptive statistics. The fever will be categorized into 4 categories;

- Grade 1: >38°C - <=39°C
- Grade 2: >39°C - <=40°C
- Grade 3: >40°C for under 24 hours
- Grade 4: >40°C for over 24 hours

The frequency of fever grade after IT section will be summarized at each visit and during the study.

8.10 Analysis of Quality of Life

Quality of life (QoL) will be analyzed in the ITT population.

8.10.1 Questionnaires

QoL will be measured using the validated questionnaires, FACT-Hep and EQ5D-3L (*Herdman 2011; Scalone 2013*), to evaluate differences in patient-reported outcomes between the 2 treatment arms.

Patients will complete the FACT-Hep, a 45-item questionnaire designed to measure Health-Related Quality of Life (HRQoL) in patients with HCC. The FACT-Hep consists of 27-item FACT-General (FACT-G), which assesses generic HRQoL concerns using 4 subscales (physical, social/family, emotional, functional wellbeing), and the 18-item hepatobiliary subscale, which evaluates specific symptoms of hepatobiliary cancer and side effects of treatment.

Patient utility will be assessed with the EQ5D-3L questionnaire.

The EQ5D-3L is composed of a descriptive system (5 domains) and a visual analog scale

(VAS). The 5 domains consist of anxiety/depression, self-care, mobility, pain/discomfort and usual activities. Each domain has 3 levels of response; no problem (coded as 1), some problems (coded as 2) and extreme problems (coded as 3). The VAS measures a person's self-rated health ranging from 0 (worst state of health imaginable) to 100 (best state of health imaginable). A health state is defined in terms of a 5-digit code. Health state code 22113, for example, would translate to some problems with mobility and self-care, no problems with performing usual activities. EQ5D-3L health states may be converted to a single index (known as a utility) by applying values (weights) to each level in each dimension. The utility is estimated by subtracting the suitable weight from 1, the value for full health (e.g. 11111).

8.10.2 Statistical Methods

The number of patients with QoL data and the number of patients missing or expected to have QoL assessments will be summarized by each treatment arm for scheduled assessment time points.

Descriptive statistics will be used to summarize the individual item domains (EQ5D-3L) and sub-scale scores (FACT-Hep) of QoL data at each scheduled assessment time point. Patients will be included if they completed at least one questionnaire item at each scheduled assessment time point. Individual item domains (EQ5D-3L) and of sub-scale scores (FACT-Hep) will be presented in listings.

9. CHANGE IN PLANNED ANALYSIS

The study was terminated early due to futility, and thus the plan for final analysis requires adjustment. Per Orphan Drug Approval Letter from the Office of Orphan Products Development/FDA (Designation #13-3942) dated 2013-May-06, the final analysis will be planned to appropriately evaluate the benefit-risk for the treatment of patients with HCC. The primary and secondary objectives will be replaced by endpoints supporting orphan drug approval as detailed in section 3.

10. REFERENCES

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.

Forner A, Ayuso C, Varela M, Rimola J, Hessheimer AJ, de Lope CR, et al. Evaluation of tumor response after locoregional therapies in hepatocellular carcinoma: are response evaluation criteria in solid tumors reliable? Cancer 2009;115:616–23.

Gordon LKK and DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika. 1983;70:659–63.

Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res.* 2011;20:1727–3.

Llovet JM, Di Bisceglie AM, Bruix J, Kramer BS, Lencioni R, Zhu A, et al. Design and endpoints of clinical trials in hepatocellular carcinoma. *JNCI* 2008;100:698–711.

O'Brien PC and Fleming TR. A multiple testing procedure for clinical trials. Biometrics. 1979;35:549–56.

Pocock SJ and Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics.* 1975;31:103–15.

Scalone L, Ciampichini R, Fagiuoli S, Gardini I, Fusco F, Gaeta L, et al. Comparing the performance of the standard EQ-5D 3L with the new version EQ-5D 5L in patients with chronic hepatic diseases. *Qual Life Res.* 2013;22:1707–16.

Yan X and Su XG. Stratified Wilson and Newcombe Confidence Intervals for Multiple Binomial Proportions. Stat Biopharm Res 2010;2;329–35.

11. APPENDICES

11.1 Tumor Response Evaluation by mRECIST

To assess tumor response, the sum of the longest diameters for all target lesions will be calculated at baseline and throughout the study. At each assessment, response is evaluated first separately for the target lesions and non-target lesions identified at baseline. These evaluations are then used to calculate the overall lesion response considering the target and non-target lesions as well as the presence or absence of new lesions.

11.1.1 Target Tumors

Target tumors show a well delineated area of viable (contrast enhancement in the arterial phase) tumor that is at least 1 cm in LD. Target tumors should be selected on the basis of their size (tumors with the LD which are able to be reproducibly measured across time points) and are preferred to be within the liver for mRECIST for HCC. However, target tumors may be selected outside of the liver if they meet criteria established by mRECIST for HCC for target tumors. When extrahepatic target tumors are selected, the contrast-enhancement characteristics are not taken into account and only LD of the tumor is measured. Up to a maximum of 10 tumors, and a maximum of 5 tumors per organ, representative of all involved organs will be identified as target tumors by the IER, and will be recorded and measured at baseline by the site reader.

All post-baseline measurements must be performed using the same tumors and methods as the baseline assessment.

Response assessments for target tumors are defined as:

- Complete Response (CR): Disappearance of all IT arterial-enhancement in all Target tumors and the disappearance of all non-enhancing Target tumors.
- Partial Response (PR): At least 30% decrease in the SLD of Target tumors, taking as reference the baseline SLD of Target tumors.
- Progressive Disease (PD): Radiographic tumor progression for target tumors requires an increase in the SLD of target tumors of at least 20% taking as reference the smallest sum of diameters of target tumors recorded since the treatment started (this includes the baseline sum if that is the smallest on study).

- Stable Disease (SD): Neither sufficient shrinkage to qualify for partial response, nor sufficient increase to qualify for progressive disease, taking as reference the smallest SLD while on study (including Baseline). Contingent upon minimum duration of 6 weeks from enrollment.
- Non-Evaluable (NE): The Reviewer is not able to evaluate Target tumors due to inadequate imaging technique or inadequate coverage. Rules for selecting non-evaluable include the following:
  1. All Target tumors are not evaluable.
  2. Or, if at least 1 Target tumor is not evaluable, the Target tumor SLD is still calculated using the remaining evaluable/measureable Target tumors. The only acceptable assessment in this situation is progressive disease or non-evaluable. If the SLD of Target tumors has increased at least 20% from nadir or the IPTP (depending on the situation) then the response is progressive disease. Any other calculated result gives an assessment of non-evaluable.

NOTE: SLD = sum of the LDs of viable enhancing hepatic target tumors plus LDs of any non-nodal extrahepatic target tumors plus the short axis diameters of any nodal target tumors) will be calculated.

11.1.2 Non-Target Tumors

Tumors with the following characteristics are defined as Non-Target Tumors by mRECIST for HCC: Infiltrative-type HCC (with ill-defined borders which are not suitable for accurate and repeat measurements) and tumors that are <1 cm in LD.

All other tumors (or sites of disease) including pathological lymph nodes should be identified as non-target tumors and should also be recorded at baseline. Measurements are not required, and these tumors should be followed as ‘present’, ‘absent’, or in rare cases ‘unequivocal progression’.

Furthermore, for the purposes of this trial, special assessments are recommended for the following:

Malignant portal vein thrombosis should be considered a non-measurable tumor due to the difficulty of performing reliable repeat measurements of a malignant thrombus.

Porta hepatis lymph node can be considered as malignant if the lymph node short axis is at least 20 mm.

Ascites, pleural effusion, and pericardial effusion: these may not be used to assess response as non-target tumors, nor may they be selected as evidence of new disease as radiographic progression. They may not be used due to the incidence of therapeutic fluid removal and benign occurrence of these fluid collections which makes them unreliable as a marker of disease evolution.

Response assessments for non-target tumors are defined as:

- Complete Response (CR): Disappearance of all non-target tumors.
- Stable Disease (SD): Persistence of 1 or more non-target tumors or IT arterial enhancement in non-target tumors.
- Progressive Disease (PD): Unequivocal progression of existing non-target tumors or the appearance of one or more new tumors (see below) is also considered progression.
- Non-Evaluable (NE): Reviewer is not able to evaluate non-target tumors due to inadequate imaging technique or inadequate coverage.

Rules for selecting non-evaluable for non-target tumors include the following:

1. All non-target tumors are not evaluable.
2. Or, if at least 1 non-target tumor is not evaluable and no other non-target tumor demonstrates unequivocal progression, the assessment is “NE”
3. If at least 1 non-target tumor is not evaluable and at least 1 other non-target tumor demonstrates unequivocal progression, the assessment is “unequivocal progression.”

11.1.3 New Tumors

New tumors in the liver are defined as being undetectable at Baseline, and subsequently becoming evident and having characteristic vascular features of HCC: arterial hypervascularization with washout in the portal venous, or the late phase of dynamic contrast imaging, if available.

New tumors outside the liver are defined as being undetectable at baseline and are subsequently clearly tumor.

New Tumor Progressive Disease (PD):

- Intrahepatic

To qualify for immediate tumor progression at the present time point (images should be sent for immediate independent confirmation of progression):

The new tumor must be ≥1 cm AND meet all of the following characteristics:

1. Hypervascular in the arterial phase
2. Area of arterial enhancement is 1 cm or larger
3. Demonstrates washout in portal- or late venous phase

Tumors that do not meet the criteria above can qualify for progression if:

The new tumor is ≥1 cm AND the grows at least 1 cm on the next sequential scan (images would be sent at this second time point for independent confirmation of progression)

- Extrahepatic

If a newly detected lesion is obviously a tumor, progression criteria is met at the current time point (images should be sent for immediate independent confirmation of progression).

The evaluation of overall response at each assessment is a composite of the target lesions response, non-target lesions response and presence of new lesions as shown in Table 2.

Table 2:
Overall Response Assessment in mRECIST; Responses for All Possible Combinations of Tumor Responses in Target and Non-target Lesions with or without the Appearance of New Lesions

Target Tumors	Non-target Tumors	New Tumors	Time Point Response
CR	CR	No	CR
CR	SD	No	PR
CR	Not evaluated (NE)	No	PR

Target Tumors	Non-target Tumors	New Tumors	Time Point Response
PR	Non-PD or NE	No	PR
SD	Non-PD or NE	No	SD
NE	Non-PD or NE	No	NE
PD	Any	Yes or No	PD
Any	PD	Yes or No	PD
Any	Any	Yes	PD

11.2 Tumor Response Evaluation by RECIST 1.1

11.2.1 Measurability of Tumor

All measurements should be recorded in metric notation (mm). At baseline, tumor lesions/lymph nodes will be categorized as measurable or non-measurable. Lymph nodes that have a short axis <10 mm at baseline are considered non-pathological and should not be recorded or followed. If no measurable lesions are identified at baseline, the patient will not be allowed to enter the study.

For tumor lesions: the longest diameter in the plane of measurement has to be recorded with a minimum size of 10 mm by CT scan when CT scan slice thickness is no greater than 5 mm.

For nodal lesions: at baseline and in the follow-up, only the short axis of lymph node will be measured and followed. To be considered pathologically enlarged and measurable, a lymph node must be >15 mm in short axis when assessed at baseline.

Non measurable lesions are defined as all other lesions present at baseline, including small lesions (longest diameter <10 mm or pathological lymph node with ≥10 mm to <15 mm short axis) as well as truly non measurable lesions.

11.2.2 Target / Non Target Tumors

Each lesion reported must be uniquely and sequentially numbered on the eCRF, even if it resides in the same organ, from baseline and throughout the study. For the evaluation of lesions at baseline and throughout the study, the lesions are classified as target and non-target lesions.

11.2.3 Target Tumors

Target tumors should be selected based on their size (tumors with the LD which are able to be reproducibly measured across time points) and are preferred to be within the liver. However, target tumors may be selected outside of the liver. Up to a maximum of 5 tumors, and a maximum of 2 tumors per organ (except for the liver where 5 tumors can be selected as non-target for the purpose of this trial), representative of all involved organs will be identified as target tumors and will be recorded and measured at baseline by the site reader.

Selection of tumors outside the liver is patient to Sponsor's approval.

All post-baseline measurements must be performed using the same tumors and methods as the baseline assessment.

Response assessments for target tumors are defined as:

- Complete Response (CR): Disappearance of all target tumors. Any pathological lymph nodes (whether target or non-target) must have reduction in short axis to <10 mm.
- Partial Response (PR): At least 30% decrease in the SLD of target tumors, taking as reference the baseline SLD of target tumors.
- Progressive Disease (PD): Radiographic tumor progression for target tumors requires an increase in the SLD of target tumors of at least 20% taking as reference the smallest sum of diameters of target tumors recorded since the treatment started (this includes the baseline sum if that is the smallest on study).
- Stable Disease (SD): Neither sufficient shrinkage to qualify for partial response, nor sufficient increase to qualify for progressive disease, taking as reference the smallest SLD while on study (including baseline). Contingent upon minimum duration of 6 weeks from enrollment.
- Non-Evaluable (NE): Progression has not been documented and one or more target lesions have not been assessed or have been assessed using a different method than baseline. Rules for selecting non-evaluable include the following:

    1. All target tumors are not evaluable.

    2. Or, if at least 1 target tumor is not evaluable, the target tumor SLD is still

calculated using the remaining evaluable/measureable target tumors. The only acceptable assessment in this situation is progressive disease or non-evaluable. If the SLD of target tumor has increased at least 20% from nadir (including baseline if it is the nadir) then the response is progressive disease. Any other calculated result gives an assessment of non-evaluable.

NOTE: SLD = sum of the LDs of viable enhancing hepatic target tumors plus LDs of any non-nodal extrahepatic target tumors plus the short axis diameters of any nodal target tumors will be calculated.

11.2.4 Non-Target Tumors

All other lesions, including pathological lymph nodes, are considered non-target lesions. Measurements of these lesions are not required and these lesions should be followed as “present”, “absent”, “worsening” or in rare cases “unequivocal progression” (as defined in the below note) throughout the study. Multiple non-target lesions involving the same organ can be assessed as a group and recorded as a single item (i.e., multiple enlarged pelvic lymph nodes). Each non-target lesion identified at baseline should be assessed at each subsequent evaluation and be recorded in the eCRF.

Furthermore, for the purposes of this trial, special assessments are recommended for the following:

Malignant portal vein thrombosis should be considered a non-measurable tumor due to the difficulty of performing reliable repeat measurements of a malignant thrombus.

Porta hepatis lymph node can be considered as malignant if the lymph node short axis is at least 20 mm.

Ascites, pleural effusion, and pericardial effusion: these may not be used to assess response as non-target tumors, nor may they be selected as evidence of new disease as radiographic progression. They may not be used due to the incidence of therapeutic fluid removal and benign occurrence of these fluid collections which makes them unreliable as a marker of disease evolution.

Response assessments for non-target tumors are defined as:

- Complete Response (CR): Disappearance of all non-target tumors. All lymph nodes must be non-pathological in size (<10 mm short axis).
- Incomplete Response / Stable Disease (SD): Neither CR nor PD

- Progressive Disease (PD): Unequivocal progression of existing non-target tumors
- Non-Evaluable (NE): Progression has not been documented and one or more non-target lesions have not been assessed or have been assessed using a different method than baseline not allowing a reliable comparison. Rules for selecting non-evaluable for non-target tumors include the following:

    1. All non-target tumors are not evaluable.
    2. Or, if at least 1 non-target tumor is not evaluable and no other non-target tumor demonstrates unequivocal progression, the assessment is “NE”
    3. If at least 1 non-target tumor is not evaluable and at least 1 other non-target tumor demonstrates unequivocal progression, the assessment is “unequivocal progression.”

NOTE: To achieve “unequivocal progression” on the basis of the non-target disease, there must be an overall level of substantially worsening in non-target disease such that, even in the presence of CR, PR or SD in target disease, the overall tumor burden has increased sufficiently to merit discontinuation of therapy. A “modest” increase in the size of one or more non-target lesions is usually not sufficient to qualify for unequivocal progression status. The designation of progression solely on the basis of change in non-target disease in the face of CR, PR or SD of target disease will therefore be extremely rare.

11.2.5 New Tumors

The appearance of new lesion is always associated with PD. A lesion identified on a follow-up assessment in an anatomical location that was not scanned at baseline is also considered a new lesion. If a new lesion is equivocal, for example because of its small size, continued therapy and follow-up evaluation will clarify if it represents truly new disease. If at the next scheduled assessment, PD is confirmed, the date of progression would be the earlier date when PD was suspected.

New Tumor Progressive Disease (PD):

If a newly detected lesion is obviously a tumor, progression criteria is met at the current time point. The evaluation of overall response at each assessment is a composite of the target lesions response, non-target lesions response and presence of new lesions as shown in Table 3.

Table 3:
Overall Response Assessment in RECIST 1.1; Responses for All Possible Combinations of Tumor Responses in Target and Non-Target Lesions with or without the Appearance of New Lesions

Target Tumors	Non-target Tumors	New Tumors	Overall Reponses
CR	CR	No	CR[1]
CR	Non-CR/non-PD	No	PR
CR	Non-PD or not all evaluated	No	PR
PR	Non-PD or not all evaluated	No	PR[1]
SD	Non-PD or not all evaluated	No	SD[1,2]
Not all evaluated	Non-PD	No	NE[1]
PD	Any	Yes or No	PD
Any	PD	Yes or No	PD
Any	Any	Yes	PD

CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease; NE, non-evaluable

[1] this overall lesion response also applies when there are no non-target lesions identified at baseline.

[2] once confirmed PR is achieved, all these assessments are considered PR.

If no non-target lesions are identified at baseline, the non-target lesion response at each assessment will be considered “not applicable” (NA).

In some circumstances, it may be difficult to distinguish residual disease from normal tissue. When the evaluation of CR depends on this determination, it is recommended that the residual lesion be investigated (fine needle aspirate / biopsy) to confirm CR. It may be sometimes reasonable to incorporate Fluorodeoxyglucose Positron Emission Tomography (FDG-PET) scanning to complement CT in assessment of progression (especially in case of possible “new” lesion) or in case where a residual radiographic abnormality is thought to represent fibrosis or scarring.

For equivocal findings of PD (e.g., very small and uncertain new lesions, cystic changes or necrosis in existing lesions) treatment may continue until the next scheduled assessment.

11.3 Re-randomization Algorithm

11.3.1 Introduction

The purpose of re-randomization is to minimize imbalance in the distribution of subjects in each treatment group within the level of prognostic factors. For each re-randomization, subjects will be reassigned to treatment groups Pexa-Vec followed by Sorafenib or Sorafenib using Pocock and Simon non-deterministic dynamic randomization algorithm, which has been used for subject randomization in the IRT system. Total 1000 randomizations will be performed independently. For the quality control perspective, the random uniform numbers U are pre-specified. A permutation test will be performed for efficacy analysis based on the 1000 re-randomizations. For each re-randomization, a log-rank test will be applied to compare the time to event of two treatment groups. Let $T_i, i = 1, \ldots, 1000$ be the test statistics for 1000 re-randomizations, and T_{obs} be the test statistic base on the observed data. An identity function $\phi(\cdot)$ is defined as,

$$\phi(T_i) = \begin{cases} 1 & if\ T_i \geq T_{obs} \\ 0 & if\ T_i < T_{obs} \end{cases}$$

The p-value of permutation test equals $\sum_{i=1}^{1000} \phi(T_i) / 1000$.

11.3.2 Dynamic Randomization Algorithm

Each patient will be re-randomized to different treatment based on minimization factor, weight, degree of imbalance, total amount of imbalance, and choice of probability of assignment.

11.3.2.1 Minimization Factor

There are six minimization factors specified in protocol. The weight and level of each minimization factor are determined by SillaJen and listed in Table 4.

Table 4 minimization factors Weight

Minimization Factor	Factor Description	Minimization Factor Weight
1	Center	1
2	Main Etiology	2
3	Extrahepatic Disease	2
4	Vascular Invasion	2
5	ECOG Performance Status	1
6	AFP Level	1

11.3.2.2 Degree of Imbalance

For each patient, the degree of imbalance $d_{ik}, i = 1, \dots, 6, and\ k = 1,2$ for each treatment groups within each minimization factor is assessed using the range of patient counts across treatment groups within the new patient's minimization factor level. For the i-th minimization factor, $i = 1, 2, \dots, 6$,

- $d_{i1} = |x_{ji1} + 1 - x_{ji2}|$, for potential assignment to Pexa-Vec followed by Sorafenib
- $d_{i2} = |x_{ji1} - x_{ji2} + 1|$, for potential assignment to Sorafenib

Where x_{ji1} is the number of subjects with level j of minimization factor i who have been assigned to treatment Pexa-Vec followed by Sorafenib and x_{ji2} is the number of subjects with level j of minimization factor i who have been assigned to treatment Sorafenib.

11.3.2.3 Total Amount of Imbalance

For each possible treatment assignment, the resulting score of total amount of imbalance G_k is the summation of weighted degree of imbalance d_{ik},

- $G_k = \sum_{i=1}^{6} w_i\, d_{ik}, k = 1, 2$

11.3.2.4 The Choices of Probability of Assignment

The probability of assignment to treatment with lowest G_k is p_1, the probability to the other treatment is $p_2 = 1 - p_1$. For SillaJen Hep024, we use fixed probability 0.6 for p_1. In the event that the G_k are tied, the probability of assignment to each treatment is 0.5.

11.3.2.5 Assign Treatment

A random number, U, is taken as the next number generated by the uniform random number generated in SAS.

- If $U \leq p_1$ then new patient is randomized to Pexa-Vec followed by Sorafenib
- If $U > p_1$and $U \leq 1$ then the new patient is randomized to Sorafenib

11.3.3 Implement the Algorithm

Step 1. A dataset PARM will be created to include all the key information for the re-randomization.

Parameter	Value
Number of Iteration	1000

Number of Treatment	2
Maximum Sample Size	600
Uniform Random Seed	1000 of Uniform seeds
Number of Minimization Factor	6
Weight of Minimization Factor	1, 2, 2, 2, 1, 1
Distance Function to Measure Imbalance	Range
Total Amount of Imbalance	Weighted Sum
Choice of Probability	0.6
Probability of Tie	1/2

Step 2. Read in minimization factors data FACTOR from clinical database

Step 3. Read in uniform randomization numbers

Step 3. Use Macro to %RERANDOM to implement the re-randomization algorithm

11.3.4 Output Result

The %RERANDOM will generate 2 output datasets: TRACE, STATISTICS.

TRACE: contains all the detail information for each re-randomization output. It contains the degree of imbalance minimization factor value, total amount of imbalance, choice of probability, and treatment assigned to subjects.

STATISTICS: contains all the statistics value per each re-randomization and P-value of the permutation test based on the formula in section 11.3.1.