A Randomized, Open-Label, International, Multi-Center, Phase 3 Clinical Study of PD-1 Antibody SHR-1210 Plus Apatinib (Rivoceranib) Mesylate Versus Sorafenib as First-Line Therapy in Subjects with Advanced Hepatocellular Carcinoma (HCC) Who Have Not Previously Received Systemic Therapy

STATISTICAL ANALYSIS PLAN

(SAP)

Protocol:	SHR-1210-III-310
Trial phase:	Phase 3
Version No.:	V 3.0
Date:	19 April, 2022

AUTHORIZED SIGNATURES

Study Title: **A Randomized, Open-Label, International, Multi-Center, Phase 3 Clinical Study of PD-1 Antibody SHR-1210 Plus Apatinib (Rivoceranib) Mesylate Versus Sorafenib as First-Line Therapy in Subjects with Advanced Hepatocellular Carcinoma (HCC) Who Have Not Previously Received Systemic Therapy**

Protocol: SHR-1210-III-310
Trial phase: Phase 3
SAP Version: V 3.0
Date: 19 April, 2022

Prepared at JHM by:

Reviewed and Approved at JHM by:

TABLE OF CONTENTS

ABBREVIATIONS

Abbreviations	**Definition**
AB	Abdominal Swelling
AE	Adverse Event
ADA	Anti-drug Antibody
AFP	Alpha-fetoprotein
AP	Appetite Loss
ALP	Alkaline Phosphatase
ALT	Alanine Aminotransferase
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Chemical Classification System
BCLC	Barcelona Clinic Liver Cancer
BI	Body Image
BID	Twice Daily
BIRC	Blinded Independent Review Committee
BOR	Best Overall Response
CF	Cognitive Functioning
CPS	Combined Positive Score
CS	Compound Symmetry
CSR	Clinical Study Report
CHMP	Committee for Medicinal Products for Human Use
CI	Confidence Interval
CMH	Cochran-Mantel-Haenszel
C_{max}	Maximum concentration
CO	Constipation
COVID	Corona Virus Disease
CPS	Combined Positive Score
CR	Complete Response
CT	Computed Tomography
CTCAE	Common Terminology Criteria for Adverse Events
D	Day
DCR	Disease Control Rate
DI	Actual Dose Intensity
DI	Diarrhoea
DILI	Drug-Induced Liver Injury
DMC	Data Monitoring Committee
DNA	Deoxyribonucleic Acid
DoR	Duration of Response
DY	Dyspnoea
eCRF	electronic Case Report Form
ECG	Electrocardiogram
ECOG-PS	Eastern Cooperative Oncology Group- Performance Status
EF	Emotional Functioning
EMA	European Medicines Agency
EORTC	European Organization for the Research and Treatment of Cancer

Abbreviations	**Definition**
EOS	End of Study
EOT	End of Treatment
EQ-5D-5L	Five-level of EuroQol-5D version
FA	Fatigue
FATI	Fatigue
FDA	Food and Drug Administration
FEV	Fever
FI	Financial Difficulties
FT	Free Triiodothyronine
GGT	γ-glutamyl Transferase
GHS	General Health Status
h	Hour
HBsAg	Hepatitis B Surface Antigen
HBV	Hepatitis B Virus
HCC	Hepatocellular Carcinoma
HCV	Hepatitis C Virus
HR	Hazard Ratio
ICH	International Council for Harmonisation
ID	Identity Document
imBOR	Immune-modified best of response
imCR	Immune-modified complete response
imNE	Immune-modified NE
imPD	Immune-modified PD
imRECIST	Immune-modified RECIST
imPR	Immune-modified partial response
imSD	Immune-modified stable disease
irAE	Immune-related Adverse Events
IRT	Interactive Response Technology
ITT	Intention- to Treat
JAUN	Jaundice
LLT	Lower Level Term
LVEF	Left Ventricular Ejection Fraction
KM	Kaplan Meier
LSMean	Least Square Mean
mRECIST	modified RECIST
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed Model Repeated Measures Model
MRI	Magnetic Resonance Imaging
Nab	Neutralizing antibody
NE	Not Estimable
NCI-CTCAE	National Cancer Institute - Common Terminology Criteria for Adverse Events
NR	Not Reached
NUTR	Nutrition
NV	Nausea and Vomiting
OBF	O'Brien-Fleming Alpha-spending Function

Abbreviations	**Definition**
OCT	Optical Coherence Tomography
ORR	Objective Response Rate
OS	Overall Survival
PA	Pain
PD	Progressive Disease
PD-1	Programmed Death-l
PFS	Progression-free Survival
PF	Physical Functioning
PK	Pharmacokinetics
PR	Partial Response
PRO	Patient Reported Outcome
RCEP	Reactive Capillary Endothelial Proliferation
RDI	Relative Dose Intensity
RECIST	Response Evaluation Criteria In Solid Tumors
RF	Role Functioning
RMST	Restricted Mean Survival Time
RNA	Ribonucleic Acid
RS	Raw Score
PT	Prothrombin Time
PT	Preferred Term
QD	Every Day
QLQ-C30	Quality of Life Questionnaire-Core 30
QLQ-HCC	Quality of Life Questionnaire-Supplement Module for HCC
QoL	Quality of life
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Stable Disease
SE	Standard Error
SF	Social Functioning
SFU	Safety Follow-Up
SL	Insomnia
SMQ	Standardized MedDRA Query
SOC	System Organ Class
SS	Safety Set
SX	Sex Life
TBIL	Total Bilirubin
TEAE	Treatment Emergent Adverse Event
TPS	Tumor Proportion Score
TNF	Tumor Necrosis Factor
TSH	Thyroid Stimulating Hormone
TL	Target Lesion
TTD	Time To Deterioration
TTP	Time To Progression
TTR	Time to Response
ULN	Upper Limit of Normal

Abbreviations	**Definition**
UI	Utility Index
UN	Unstructured
VAS	Visual Analogue Scale

AMENDMENT HISTORY

Version	Version Date	Summary of Changes
1.0	30-Sep-2020	Not applicable (N/A)
2.0	05-Aug-2021	Section 3.1.1 Overall Survival and 3.1.2 Progression-free survival by BIRC assessment per RECIST v1.1: Updated the section. Section 3.3 Safety Endpoints: Clarified the definition of on- treatment period. Section 3.3.1 Adverse Events: Provided further details for the selected adverse events defined by sponsor. Deleted the definitions for time to first onset of irAE and time to irAE resolution. Deleted the liver function test abnormalities. Section 3.3.2 Laboratory Test: Added the definition of potential Hy's law. Section 3.3.3 Vital Signs: Updated clinically significant abnormal criteria for Systolic Blood Pressure and Diastolic Blood Pressure in Table 5. Section 3.3.4 12-lead ECG: Clarified the ECG measurements and blood sample collected time. Section 3.3.6 ECOG-PS score: Added the section. Section 3.3.7 Child-Pugh: Added the section. Section 3.3.8 Echocardiography: Added the section. Section 3.3.9 Treatment Exposure: Added details of treatment exposure. Section 3.4 Pharmacokinetic and Immunogenicity and section 4.14 "Pharmacokinetics and Immunogenicity Endpoint Analyses": Added the efficacy endpoints and safety endpoints, added definitions for Nab and ADA and analyses for ADA and Nab. Section 3.5.1.1 EORTC QLQ-C30: Added definition of TTD for the symptom's scales/items. Clarified the TTD definition for HRQoL/function. Section 3.5.3 Biomarker Endpoints and the Other Endpoints: Added the categories for expression level. Section 4.2 Definition of Study Window and Baseline: Modified the visit window and target days in Table 12. Section 4.3.7 PK: Updated the rule for PK summary. Section 4.3.7 Analysis Populations: Added the PKQTS population. Section 4.5 Protocol Deviations: Deleted the non-important protocol deviations and Added the summary for all protocol deviations related to COVID-19. Section 4.7 Demographics and Baseline Characteristics: Updated the section. Section 4.9 Prior and Subsequent Anti-Cancer Therapy: Added the definitions of prior and subsequent anti-cancer therapy. Section 4.11 Efficacy Analyses: Updated the section.

		Section 4.13.1 Patient Reported Outcome (PRO) Endpoints Analyses: Added the summary for complete rates of PRO endpoints and updated the analysis method. Section 4.14.1 Pharmacokinetics: Rephrased the whole section for clarification. Section 4.14.2 Immunogenicity Analysis of SHR-1210: Modified the definition of ADA/Nab positive and negative, time to and duration of ADA/Nab response. Added the sections of Demographics and Baseline Characteristics, Impact of Immunogenicity on PK of SHR-1210, Impact of Immunogenicity on Efficacy of SHR-1210 and Safety of SHR-1210 for Analyses. Section 4.15 Safety Analyses: Added the summary for missing visits due to COVID-19 and discontinuation or interruption of study drugs due to COVID-19. Section 4.15.2 All Adverse Events: Added the overview of summary of immune-related TEAEs assessed by investigator and COVID-19 related TEAEs. Added some summary tables of AEs. Deleted the summary of TEAEs leading to interruption of all study drug. Some analyses were combined in one table. Section 4.15.2.1.2 irAEs assessed by investigator: Deleted the analyses of irAEs including time to first onset of irAE, resolution of irAEs and duration of systemically administered corticosteroids. Section 4.15.2.1.3 Selected Adverse Events: Added the summaries of SHR-1210 selected adverse events and Rivoceranib selected adverse events and time to first onset of selected AEs, resolution of selected AEs， and duration of systemically administered corticosteroids. Section 4.15.2.1.4 COVID-19 related TEAEs: Added the summary of COVID-19 related TEAEs leading to any study drug interruption or dose reduced by SOC and PT. Section 4.15.3 Laboratory Tests Results: Added analyses of urinalysis, potential Hy's law and listing of potential Hy's law. Section 4.15.5 Child-Pugh: Added the analysis of Child-Pugh. Section 4.15.7 Physical Examination: Moved to section 4.15.7. Section 4.15.9 Concentration-QTc analysis for Rivoceranib: Added the analysis. Section 4.15.10 Echocardiography: Added the section. Section 4.18 Subgroup Analyses: Added the subgroup analyses of ORR assessed by BIRC, added PD-L1 categories and modified some subgroup categories. Added the analyses of demographic and other baseline characteristics by geographical region. Minor editorial and consistency changes throughout the document.
3.0	19-Apr-2022	Section 1 Introduction: Updated the protocol version to 5.1 Section 3.2.6 Time to Response (TTR) by BIRC or Investigator: Added the section Section 3.2.7 PFS, ORR, DCR, DoR, TTP， TTR Evaluated by BIRC based on Modified RECIST (mRECIST): Added definition for TTR

		Section 3.3.1 Adverse Events: Added an example for the time to resolution of selected AEs. Added the definition for duration of systemically administered corticosteroids. Section 3.3.2 Laboratory Test: Updated the definition of the concurrent measurement. Section 3.5.2 imRECIST Endpoints Evaluated by Investigator: Added the definition for immune-modified TTR (imTTR) evaluated by investigator Section 4.3.2 Adverse Events: Updated the imputation rule for missing end dates. Section 4.3.3 Concomitant Medications: Updated the imputation rule for missing end dates. Section 4.7 Demographics and Baseline Characteristics: Added non-viral into the root cause of HCC. Section 4.11.2.6 Time to Response: Added the section. Section 4.15.2.1.1 All Adverse Events: Added the exposure adjusted incidence rate analyses for the overall TEAE, TEAE with CTCAE grade ⩾3, TEAE excluding hypertension, and TEAE excluding hypertension with CTCAE grade ⩾3. Section 4.15.2.1.2 irAEs Assessed by Investigator: Added the analyses for number and percentage of subjects with at least one irAE treated with any hormone replacement therapy for overall, duration (total days) of systemically administered Corticosteroids by medical concept category for subjects treated with ⩾ 40mg/day systemically administered Corticosteroids, duration (total days) of systemically administered Corticosteroids ⩾ 40mg/day by medical concept category, number and percentage of subjects with at least one selected AEs thyroid disorder and type 1 diabetes Mellitus for SHR-1210 treated with any hormone replacement therapy by medical concept category and by maximum CTCAE grade, treatment of Selected AEs hepatotoxicity for Rivoceranib. Section 4.15.3 Laboratory Test Results: Updated the definition of concurrent measurements. Section 4.18 Subgroup Analysis: Updated that ORR was based on BIRC; Updated that disease organs was based on BIRC, the definition for the previous local therapy. Added OS/PFS analysis for Mainland China/HongKong/Taiwan. Appendix 1. Selected Adverse Events defined by Sponsor: Updated the category and SMQs/PTs. Appendix 3. SAS code for primary analysis of PFS and OS: Added this section. Minor editorial and consistency changes throughout the document.

1. INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to describe the statistical methods, data derivations and data presentation to be performed on this study before any formal statistical analysis or database locked, and to ensure the credibility of the study findings by further elaborating details regarding the definition of analysis variables and analysis methodology as pre-specified in study protocol to address all study objectives. The SAP will serve as a compliment to the protocol, and where there is a discrepancy between the SAP and protocol, the SAP supersedes it. Any critical revision within the SAP of definition(s) or analysis method(s) for the primary endpoints will be included in a protocol amendment.

The preparation of this SAP has been developed in conjunction with the International Conference on Harmonization (ICH) relevant guidelines, European Medicines Agency (EMA/CHMP Note for Guidance on Choice of Control Group in Clinical Trials 2001), FDA (Guidance for Industry: Adaptive Designs for Clinical Trials of Drugs and Biologics 2018) and study protocol (SHR-1210-III-310 V5.1).

1.1. Study Design

This is a randomized, open-label, international, multi-center, phase 3 trial to evaluate the efficacy and safety of PD-1 antibody SHR-1210 plus rivoceranib mesylate versus sorafenib as first-line therapy in subjects with incurable advanced HCC.

The study will be conducted in subjects with incurable, locally advanced or metastatic HCC who did not receive systematic treatment previously. Primary efficacy endpoints include both overall survival (OS) and progression free survival (PFS) evaluated by BIRC based on RECIST v1.1, and approximately 510 subjects will be enrolled. Eligible subjects will be randomized to receive either SHR-1210 combined with rivoceranib mesylate (experimental arm) or sorafenib (control arm) in a 1:1 ratio. The stratification factors for randomization include:

1. Macrovascular invasion and/or extrahepatic metastasis (presence vs. absence);

2. Geographical region (Asia vs. countries outside of Asia);

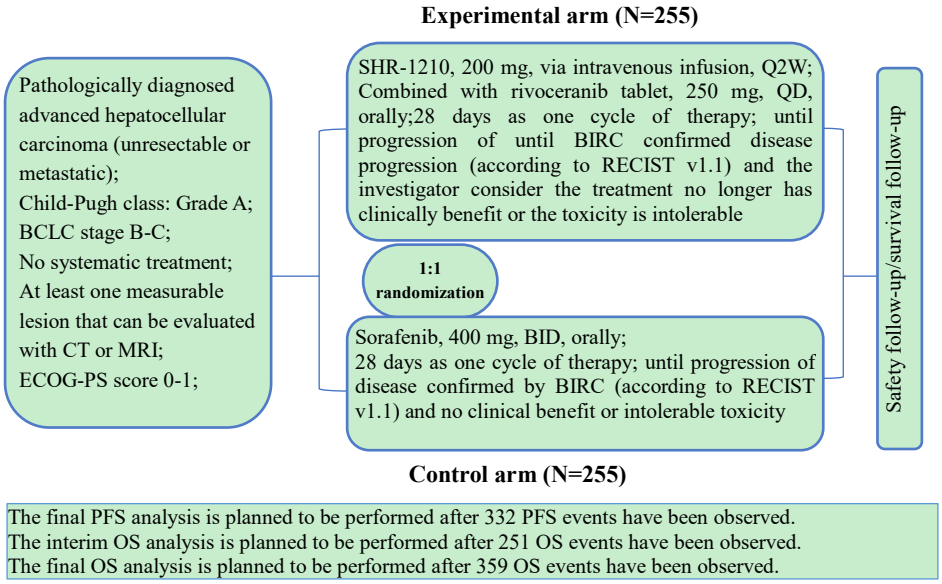
3. Baseline AFP (AFP < 400 ng/mL vs. AFP ⩾400 ng/mL).

Subjects will receive study treatment after being informed of all pertinent aspects of the study, signing the informed consent form and passing the screening for eligibility. Experimental arm: SHR-1210, 200 mg, via intravenous infusion, once every two weeks (Q2W) + rivoceranib 250 mg, p.o., once per day (QD), continuously, 4 weeks (28 days) per cycle of therapy. Control arm: sorafenib, 400 mg, p.o., twice per day (BID), continuously, 28 days per cycle of therapy. Study treatment will continue until the subject develops an intolerable toxicity, withdrawing informed consent, and disease progression confirmed by BIRC according to RECIST v1.1 (when the subject has disease progression assessed by the investigator according to RECIST v1.1, the investigator must submit the radiological data to BIRC immediately. If BIRC evaluates it as non-disease progression according to RECIST v1.1, the subject should continue to receive the study drug treatment and continue the tumor radiological

evaluation; if BIRC confirms it as disease progression based on RECIST V1.1, at this time, the investigator needs to assess whether the subject still has clinical benefit. If the subject is considered to still have clinical benefit and meet the criteria for treatment beyond disease progresses, the subject may continue to receive study medication; if the subject is no longer considered to have clinical benefit, the treatment may be discontinued), or other termination criteria specified by the protocol, whichever occurs first.

Subjects will have safety visits on D1 and D15 of each cycle of therapy within the first 3 cycles for both experimental and control arms. Beginning from the cycle 4, subjects will have safety visits on D1 and D15 of each cycle in the experimental arm; subjects in the control arm will only have safety visits on D1 of each cycle. Subjects will continue with safety visits and survival follow-up after the end- of- treatment. The study design is shown below:

Study Design Chart



The final PFS analysis is planned to be performed after 332 PFS events have been observed.
The interim OS analysis is planned to be performed after 251 OS events have been observed.
The final OS analysis is planned to be performed after 359 OS events have been observed.

Tumor radiological evaluation will be performed every 8 weeks (56 days) ± 7 days from the randomization through the first 48 weeks, to assess the efficacy, and every 12 weeks (84 days) ± 7 days after 48 weeks. During the trial period, radiological examination and evaluation can be added at any time if clinically indicated. Tumor radiological assessment will continue until the occurrence of disease progression confirmed by BIRC according to RECIST v1.1 criteria or study treatment discontinuation, whichever occurs later. Subjects who discontinue treatment for reasons other than BIRC-confirmed disease progression (according to RECIST v1.1) will also continue with regular follow-up by tumor radiological assessments after study treatment is discontinued.

If the subject withdraws informed consent before BIRC confirms disease progression according to the RECIST v1.1 criteria or the study treatment is discontinued, other anti-tumor treatment has begun (except for traditional Chinese medicine), or the subject dies, there is no need to continue the radiological evaluation. If the subject does not meet the above criteria to stop radiological assessment, the efficacy evaluation of the three efficacy evaluation criteria (RECIST v1.1, mRECIST,

imRECIST) needs to continue even if the disease progression under certain efficacy evaluation criteria occurs.

1.2. Objectives and Statistical Hypotheses

1.2.1. Primary Objectives

To compare efficacy endpoints of OS and PFS evaluated by BIRC based on RECIST v1.1 in 2 treatment groups (SHR-1210 in combination with rivoceranib versus sorafenib).

1.2.2. Key Secondary Objectives

To compare the ORR (assessed by BIRC according to RECIST v1.1) of SHR-1210 in combination with rivoceranib to sorafenib

1.3. Sample Size Determination

Assuming a 1:1 randomization between sorafenib and SHR-1210+rivoceranib, and a median PFS time of 3.6 months for subjects in the sorafenib group, and 6 months in SHR-1210 group, 332 PFS events will be sufficient to detect a hazard ratio (HR) of 0.60 with 98% power based on Log-rank test at significance level of 0.005 one-sided. Similarly, assuming a median survival time of 10.5 months for subjects in the sorafenib group, and 14.6 months in SHR-1210 group, 359 OS events will be sufficient to detect a hazard ratio (HR) of 0.72 with 85% power based on Log-rank test at significance level of 0.020 one-sided. Assuming an accrual period of 18 months, total duration of 36 months and 19 months for the OS and PFS, respectively, and probability of subject loss be approximately 10% per 12 months and per 6 months for the OS and PFS, respectively, a sample size of 510 patients will be enrolled for the study. Power and sample calculations were performed using EAST® 6.4.1. software package.

2. STATISTICAL HYPOTHESES AND DECISION RULES

The following statistical hypotheses will be tested to address the primary objectives:

H_{01}: $HR_{PFS} \geq 1$ vs H_{11}: $HR_{PFS} < 1$

H_{02}: $HR_{OS} \geq 1$ vs H_{12}: $HR_{OS} < 1$

Where HR_{PFS} and HR_{OS} are the hazard ratios of PFS and OS respectively.

The overall type I error is controlled at 0.025 1-sided by allocating α_1=0.005 to the PFS comparison and by allocating α_2=0.020 to the OS comparison.

Superiority of SHR-1210 plus rivoceranib over sorafenib in PFS is demonstrated, if one-sided p-value obtained from the stratified log-rank test is significant at 0.005 or 0.025 (if H_{02} is rejected) in favoring the experimental group.

Superiority of SHR-1210 plus rivoceranib over sorafenib in OS is demonstrated, if

- one-sided p-value obtained from the stratified log-rank test is significant at 0.005 at interim analysis (~70% events observed), or at 0.018 at final analysis in favoring the experimental group;

or

- one-sided p-value obtained from the stratified log-rank test is significant at 0.007 at the interim analysis (~70% events observed), or at 0.023 at final analysis (if H_{01} is rejected) in favoring the experimental group.

The significance levels for interim and final efficacy analyses for OS will be determined by using the Lan-DeMets procedure with an O'Brien-Fleming stopping boundary based on actual number of deaths observed.

The following statistical hypothesis will be tested to address the key secondary objective:

H_{03}: Difference of ORR$\leqslant$0 vs H_{13}: Difference of ORR$>$0

If both the PFS and OS endpoints are significant, ORR assessed by BIRC according to RECIST v1.1 at the time of 70% OS events observed will be tested at 0.025 one-sided for SHR-1210 plus rivoceranib compared to sorafenib using stratified Cochran-Mantel-Haenszel (CMH) test. If both PFS and OS endpoints are not significant, then the testing for ORR will not be performed.

Details of multiplicity control for the comparisons for PFS, OS, and ORR can be found in Section 4.17.

3. STUDY ENDPOINTS

3.1. Primary Endpoints

3.1.1. Overall Survival

Overall Survival (OS): defined as time from randomization to death due to any cause (i.e., date of death or censoring – date of randomization +1). Subjects without documented death at the time of analysis will be censored based on the last recorded date on which the subjects were known to be alive.

For any OS analysis, it may be necessary to use all relevant CRF fields to determine the last recorded date on which the subject was known to be alive. The last date for each individual subject is defined as the latest among the following dates recorded on the eCRFs：

- Randomization date
- Start treatment and end of treatment date
- Start and end date of concomitant therapy
- AE start and end dates

- All subject assessment dates (Laboratory test [laboratory, pharmacokinetic], vital signs, 12-lead ECG, echocardiography, physical examination, Child-Pugh score, ECOG-PS, tumor assessments, subject's self-evaluation)
- Date of last dose on 'end of treatment' eCRF pages
- Start and end dates of anti-cancer therapies administered after study treatment discontinuation
- Date last known alive on the 'Survival Follow-up' CRF (do not use date of survival follow-up assessment unless status is 'Alive')
- End of study date (do not use the date of death in the page)

During the interim analyses, for subjects who are confirmed to be alive or die after the data cut-off date, they will be censored at the cut-off date.

Reasons for censoring will be derived in Table 1 following the hierarchy shown below:

Table 1. Censoring Reasons for OS analyses

No.	Condition	Censoring Reason
1	No event and [reason of End of study (EOS)=Withdrawal by subject]	Withdrawal of consent
2	No event and [EOS=Lost to Follow-up/Physician decision in EOS page]	Lost to follow-up
3	No event and none of the conditions in the above hierarchy is met	Alive

3.1.2. Progression-Free Survival by BIRC Assessment per RECIST v1.1

Progression-free-survival (PFS): defined as the time from the date of randomization to the first occurrence of radiological progression by tumor image evaluation or death, whichever comes first, evaluated by BIRC based on RECIST v1.1, regardless of whether the subject withdraws from randomized therapy.

PFS data will be censored on the date of the last adequate tumor assessment for subjects who do not have an event (PD or death), for subjects who start a new anti-cancer therapy (except for traditional Chinese medicine) prior to an event or for subjects with an event after 2 or more missing tumor assessments. An adequate post-baseline assessment is defined as an assessment where a response of CR, PR, SD, non-CR/non-PD, or PD can be determined. Time points where the response is NE or no assessment was performed will not be used for determining the censoring date. Subjects who do not have a baseline or adequate post-baseline tumor assessment will be censored on the date of randomization unless death occurred on or before the time of the planned tumor assessment (i.e., ≤16weeks after the date of randomization) in which case the death will be considered as an event.

Given the scheduled visit assessment schedule (i.e., eight-weekly for the first 48 weeks then twelve-weekly thereafter) the definition of 2 missed visits is defined as below:

- If the previous tumor assessment is conducted prior to study day 273 (i.e., week 39) then two missing visits will equate to 17 weeks after the previous tumor assessment, allowing for early and late visits.

- If the two missed visits occur over the period when the scheduled frequency of tumor assessments changes from eight-weekly to twelve-weekly this will equate to 21 weeks. The time period for the previous tumor assessment will be from study days 273 to 343 (i.e., week 39 to week 49).

- From week 49 onwards (when the scheduling changes to twelve-weekly assessments), two missing visits will equate to 25 weeks.

The PFS time will always be derived based on scan dates not visit dates.

In the instance where there are different dates of scans within the same tumor assessment, the response assessment will use the last scan date where lesions are defined and an assessment of PD by investigator will be dated on the earliest scan date that demonstrates PD assessed by investigator:

- The date of the first scan (from baseline) on which the New Lesion was detected (if a New Lesion was detected at that visit)

- The date of the first scan (from baseline) of Target Lesions (if the Target Lesion response is progression)

- The date of the first scan (from baseline) on which the Non-Target Lesion(s) progressed (if the Non-Target Lesion response is progression).

The censoring and event date options to be considered for the PFS analyses are presented in Table 2.

PFS (months)= [date of PFS event or censoring – date of randomization + 1]/30.4375

Table 2. Censoring Rules for PFS Analyses

Scenario	Date of censoring
No baseline or no adequate post-baseline disease assessment	Censored at date of randomization
No PD, no death and no new anticancer treatment initiated	Censored at last adequate tumor assessment
No PD, no death and new anticancer treatment initiated	Censored at last adequate tumor assessment before new anti-cancer treatment
PD or death after 2 or more missing or inadequate post-baseline tumor assessments	Censored at last adequate tumor assessment prior to the ≥ 2 missed or inadequate post-baseline tumor assessment

Reasons for censoring in hierarchical order are shown in Table 3:

Table 3. Censoring Reasons for PFS Analyses

No.	Condition	Censoring Reason
1	No baseline assessment	No baseline assessment
2	Start of new anti-cancer therapy	Start of new anti-cancer therapy
3	PD/Death after 2 or more missing or inadequate post-baseline tumor assessments	Event after 2 or more missing assessments [a]
4	No PD/Death and Reason of EOS=Withdrawal by subject	Withdrawal of consent
5	No PD/Death and Reason of EOS= Lost to follow-up	Lost to follow-up
6	No PD/Death and [end of study (EOS) page present] and no adequate post-baseline tumor assessment	No adequate post-baseline tumor assessment
7	No PD/Death and none of the conditions in the above hierarchy are met	Ongoing without an event

[a] 2 or more missing or inadequate post-baseline tumor assessments.

3.2. Secondary Endpoints

3.2.1. Progression-Free Survival by Investigator Assessment per RECIST v1.1

Progression-free-survival (PFS): defined as the time from the date of randomization to the first occurrence of radiological progression by tumor image evaluation or death, whichever comes first, evaluated by the investigator based on RECIST v1.1. regardless of whether the subject withdraws from randomized therapy.

The censoring rule and other rules are the same as PFS by BIRC assessment per RECIST v1.1.

3.2.2. Objective Response Rate (ORR) by BIRC or Investigator Assessment per RECIST v1.1

ORR as evaluated by BIRC based on RECIST v1.1: defined as the percentage of subjects with complete response (CR) or partial response (PR) evaluated by BIRC based on RECIST v1.1.

ORR as evaluated by investigator based on RECIST v1.1: defined as the percentage of subjects with complete response (CR) or partial response (PR) evaluated by investigator based on RECIST v1.1.

The assessment of ORR will include all assessments obtained up until progression, or the last assessment in the absence of progression prior to starting new anti-cancer therapy.

Best Objective Response

Best objective response (BOR) is calculated based on the overall visit response from each RECIST assessment. It is the best response a subject has had following randomization but prior to starting any new anti-cancer therapy and prior to RECIST progression or the last assessment in the absence of RECIST progression and new anti-cancer therapy.

Categorization of confirmed BOR will be based on RECIST using the following response categories: CR, PR, SD, PD and NE.

- CR: at least two determinations of CR at least 4 weeks apart and before first documentation of PD. It is reasonable to consider a subject with time point responses of CR-NE-CR, CR-NE-NE-CR as a confirmed complete response.

- PR: at least two determinations of PR or better (PR followed by PR or PR followed by CR) at least 4 weeks apart and before first documentation of PD (and no qualifying for a CR). It is reasonable to consider a subject with one NE or two NEs, or one SD occurred between the two response as a confirmed partial response.

- SD: at least one SD assessment (or better) $\geq$ 8 weeks $\pm$1 week after the date of randomization and before first documentation of PD (and not qualifying for CR or PR).

- PD: first documentation of PD $\leq$ 17weeks after the date of randomization (and not qualifying for CR, PR, SD).

- NE: all other cases.

BOR will be determined programmatically based on RECIST from the overall visit response at each visit using BIRC data including all assessments obtained up until the first progression, the start of any new anti-cancer therapy or the last assessment in the absence of progression. For subjects whose PFS event is death, BOR will be calculated based upon all RECIST assessments prior to death.

For a best response of SD, the latest of the dates of contributing towards a particular overall visit assessment will be used. SD should be recorded at least 8 weeks $\pm$1 week after randomization. For CR/PR, the initial overall visit assessment which showed a response will use the latest the dates contributing towards a particular overall visit assessment.

Change in Tumor Size

The percentage change from baseline in tumor size will be derived at each scheduled tumor assessment visit (i.e., week 8, week 16 etc. hereafter referred to as week X for convenience). Best percentage change from baseline in tumor size will be derived as the biggest decrease or the smallest increase in tumor size from baseline until documented disease progression, excluding assessments after start of new anti-cancer therapy.

This is based on RECIST target lesion (TL) measurements taken at baseline and at the time point of interest. Tumor size is defined as the sum of the longest diameters of the

TLs based upon RECIST assessments. TLs are measurable tumor lesions. Baseline for RECIST is defined to be the last evaluable assessment prior to randomization. The change in tumor size at week X will be obtained for each subject by taking the difference between the sum of TLs at week X and the sum of the TLs at baseline. To obtain the percentage change in TL tumor size at week X the change in tumor size is divided by the sum of TLs at baseline and multiplied by 100 (i.e. (week X – baseline) / baseline *100).

Whenever tumor size data for the week X visit (Note: or visit at which progression was documented if before week X) is available then this should be used in the analysis. A windowing rule will be applied and will follow the protocol allowed visit window; Therefore any RECIST scan performed within $\pm$ 1 week of the protocol visit will be used for that visit.

Measurement from the reviewer selected by adjudicator will be used when adjudication for overall visit response has occurred, but in the case where no adjudication was required, the measurements from the reviewer who was assigned as reviewer 1 in Lesion Management Solutionwill be used for this analysis.

3.2.3. Disease Control Rate (DCR) by BIRC or Investigator Assessment per RECIST v1.1

DCR as evaluated by BIRC based on RECIST v1.1: defined as the percentage of subjects with complete response, partial response or stable disease (SD) $\geq$ 8 weeks $\pm$1 week evaluated by BIRC based on RECIST v1.1.

DCR as evaluated by investigator based on RECIST v1.1: defined as the percentage of subjects with complete response, partial response or stable disease (SD) $\geq$ 8 weeks $\pm$1 week evaluated by investigator based on RECIST v1.1.

3.2.4. Duration of Response (DoR) by BIRC or Investigator Assessment per RECIST v1.1

DoR as evaluated by BIRC based on RECIST v1.1: defined as time from the date of first record of objective response (CR or PR) to the first occurrence of radiological progression or death, whichever comes first, evaluated by BIRC based on RECIST v1.1.

DoR as evaluated by investigator based on RECIST v1.1: defined as time from the date of first record of objective response (CR or PR) to the first occurrence of radiological progression or death, whichever comes first, evaluated by investigator based on RECIST v1.1.

DoR (months) = [date of event or censoring– first date of CR/PR +1]/30.4375

The time of the initial response will be defined as the latest of the dates contributing towards the first visit response of CR or PR.

If a subject does not progress following a response, then the corresponding DoR will be censored at the PFS censoring time.

DoR will not be defined for those subjects who do not have documented response.

3.2.5. Time to Progression (TTP) by BIRC or Investigator Assessment per RECIST v1.1

TTP as evaluated by BIRC based on RECIST v1.1: defined as the time from the date of randomization to radiological progression of disease evaluated by BIRC based on RECIST v1.1.

TTP as evaluated by investigator based on RECIST v1.1: defined as the time from the date of randomization to radiological progression of disease evaluated by investigator based on RECIST v1.1.

TTP (months)= [date of TTP event or censoring – date of randomization + 1]/30.4375

Table 4. Censoring Rules for TTP Analyses

Scenario	Date of censoring
No baseline or no adequate post-baseline disease assessment	Censored at date of randomization
No PD and no new anti-cancer treatment initiated	Censored at last adequate tumor assessment
No PD and new anti-cancer treatment initiated	Censored at last adequate tumor assessment before new anti-cancer treatment is given
No PD but death before 2 missing or inadequate post-baseline tumor assessments and with no new anti-cancer treatment	Censored at last adequate tumor assessment
PD/death after 2 or more missing or inadequate post-baseline tumor assessments	Censored at last adequate tumor assessment prior to the ≥ 2 missed or inadequate post-baseline tumor assessment

Reasons for censoring in hierarchical order are shown in Table 5:

Table 5. Censoring Reasons for TTP Analyses

No.	Condition	Censoring Reason
1	No baseline assessment	No baseline assessment
2	Start of new anti-cancer therapy	Start of new anti-cancer therapy
3	PD/Death after 2 or more missing or inadequate post-baseline tumor assessments	PD/Death after 2 or more missing assessments [a]
4	No PD and Reason of EOS=Withdrawal by subject	Withdrawal of consent
5	No PD and Reason of EOS= Lost to follow-up	Lost to follow-up
6	Death	Death
7	No PD and [end of study (EOS) page present] and no adequate post-baseline tumor assessment	No adequate post-baseline tumor assessment
8	No PD and none of the conditions in the above hierarchy are met	Ongoing without PD

3.2.6. Time to Response (TTR) by BIRC or Investigator Assessment per RECIST v1.1

TTR as evaluated by BIRC based on RECIST v1.1: defined as the time from the date of randomization to the date of first record of objective response (CR or PR) evaluated by BIRC based on RECIST v1.1.

TTR as evaluated by investigator based on RECIST v1.1: defined as the time from the date of randomization to the date of first record of objective response (CR or PR) evaluated by investigator based on RECIST v1.1.

TTR (months)= [first date of CR/PR– date of randomization + 1]/30.4375

3.2.7. PFS, ORR, DCR, DoR, TTP, TTR Evaluated by BIRC based on Modified RECIST (mRECIST)

PFS as Evaluated by BIRC based on Modified RECIST (mRECIST): defined as the time from the date of randomization to the first occurrence of radiological progression by tumor image evaluation or death, whichever comes first, evaluated by BIRC based on mRECIST.

ORR as Evaluated by BIRC based on Modified RECIST (mRECIST): defined as the percentage of subjects with complete response (CR) or partial response (PR) evaluated by BIRC based on mRECIST.

DCR as Evaluated by BIRC based on Modified RECIST (mRECIST): defined as the percentage of subjects with complete response, partial response or stable disease (SD) $\geq$8 weeks $\pm$1 week evaluated by BIRC based on mRECIST.

DoR as Evaluated by BIRC based on Modified RECIST (mRECIST): defined as time from the date of first record of objective response (CR or PR) to the first occurrence of radiological progression or death, whichever comes first, evaluated by BIRC based on mRECIST.

TTP as Evaluated by BIRC based on Modified RECIST (mRECIST): defined as the time from the date of randomization to radiological progression of disease evaluated by BIRC based on mRECIST.

TTR as Evaluated by BIRC based on Modified RECIST (mRECIST): defined as the time from the date of randomization to the date of first record of objective response (CR or PR) evaluated by BIRC based on mRECIST.

The censoring rules and other rules for PFS, DoR and TTP by BIRC based on mRECIST are the same as PFS, DoR and TTP assessed by BIRC per RECISIT.

3.3. Safety Endpoints

Safety data collected during the on-treatment period will be summarized unless specified otherwise. “on-treatment” is defined as the time from the first dose through minimum (start day of first new anti-cancer therapy after last dose date-1 day, 90 days

+ last dose of study drug). Safety data collected outside the on-treatment period will not be summarized unless specified otherwise, but will be listed and flagged.

3.3.1. Adverse Events

An adverse event (AE) can be any unfavorable and unintended symptom, sign, abnormal laboratory finding, or disease, including the worsening of pre-existing medical conditions/diseases, any new adverse medical conditions, or abnormal clinically significant laboratory values or results that are not caused by concomitant disease. For reporting purposes, protocol defined that collection of AE/SAE begins from the time of subject informed consent, through and including 90 days after the last dose of SHR-1210 or 30 days after the last dose of targeted anti-angiogenic therapy, whichever comes later. All serious AEs suspected of study drug related occur after this period are also collected. All AEs/SAEs are to be followed up until resolve, return to baseline level or ≤ Grade 1 (CTCAE v4.03), reach a stable state, or until loss of follow-up or death. All AEs will be coded using a single MedDRA dictionary (v 24 or higher) for preferred term (PT) and primary system organ class (SOC). Severity of AE will be graded using the NCI CTCAE v4.03 criteria. For the purposes of analysis, a treatment-emergent adverse event (TEAE) is an AE with an onset date during the on-treatment period.

- **Treatment-Related Adverse Events:** adverse events with relationship to study treatment (as recorded on "Adverse Event" eCRF page, Relationship to study treatment= "Related"). Adverse events that have missing causality will be considered as related to study drug.

- **Adverse Events Leading to Dose Reduction:** adverse events leading to dose reduction of study treatment (as recorded on "Adverse Event" eCRF page, Action to study treatment= "Dose Reduced or Dose interrupted and Reduced"). Note: no dose adjustment is allowed for SHR-1210 treatment component;

- **Adverse Events Leading to Interruption of Study Drug:** adverse events leading to interruption of study treatment (as recorded on "Adverse Event" eCRF page, Action to study treatment= ("Dose interrupted" or "Dose interrupted and Reduced");

- **Adverse Events Leading to Both Dose Reduction and Interruption of Any Study Drug:** adverse events leading to interruption of study treatment (as recorded on "Adverse Event" eCRF page, Action to study treatment= (Dose interrupted and Reduced)). Note: no dose adjustment is allowed for SHR-1210 treatment component;

- **Adverse Events Leading to Discontinuation of Study Drug:** adverse events leading to treatment discontinuation (as recorded on "Adverse Event" eCRF page, Action to study treatment= "Drug Withdrawn").

Selected Adverse Events Defined by Sponsor：

- **Selected Adverse Events for SHR-1210**

For the purposes of analysis, a set of comprehensive definitions using standardized

MedDRA queries (SMQs), and Sponsor-defined AEGTs (adverse event grouped terms) were adopted to identify selected Adverse Events by medical concept from the adverse event clinical database. These medical concepts included SHR-1210 associated important identified and potential risks, and class effects reported with other immune-checkpoint inhibitors, or important identified risks associated with the use of rivoceranib.

A complete list of Selected AEs for SHR-1210 is described in Appendix 1.

Selected AEs Treated with Systemic Corticosteroids for SHR-1210
These events are a subset of the SHR-1210 Selected AEs and were determined by their temporal relationship with the use of systemic corticosteroids to elucidate the possible association of these events and the body's humoral or cell-mediated immunity. Data collected on the concomitant medication page and AE page of the eCRF were both used to programmatically identify the selected AEs requiring the use of systemic corticosteroids.

Selected AEs using systemic corticosteroid for SHR-1210 were generated and include AEs according to the following criteria:

- systemic corticosteroid was given between the AE onset date and the AE resolution date.

Note: The AEs with missing resolution dates were included. The AEs with a resolution date same as the systemic corticosteroid initiation date were not included.
Systemic corticosteroid was identified via the ATC codes as show below. Only medications with the route of "intramuscular", "intravenous", "subcutaneous", "oral" or missing are considered. Medications with the route of "topical" or "other" are considered as non-systemic.

Systemic corticosteroids:
H02A-corticosteroids for systemic use, plain
H02B- corticosteroids for systemic use, combination

Selected AEs Requiring Immunosuppressive Treatment Other Than Corticosteroids for SHR-1210
Immunosuppressive agents other than corticosteroids were identified via ATC codes:
L04AA- selected immunosuppressants
L04AB-Tumor Necrosis Factor alpha (TNF-α) inhibitors
L04AC-Interleukin inhibitors
L04AD-Calcineurin Inhibitors
L04AX-other immunosuppressants

Selected AEs Requiring Hormone replacement therapy for SHR-1210

Hormone replacement for categories of Thyroid disorder and Type 1 Diabetes Mellitus will be identified via the ATC codes as shown below:
H03A-thyroid preparations
H03B-antithyroid preparations
A10A-insulins and analogues

- **Selected Adverse Events for Rivoceranib**

To further characterize the rivoceranib toxicities, selected AEs for rivoceranib were also summarized by medical concepts, using MedDRA-standardized SMQs, or Sponsor-defined AEGTs.

A complete list of selected AEs for rivoceranib is described in Appendix 1.

Selected AEs hepatotoxicity for Rivoceranib Treated with Systemic Corticosteroids and Requiring Immunosuppressive Treatment Other Than Corticosteroids will be identified as selected AEs for SHR-1210.

Time to First onset of Selected AEs:

Time (Months) to first onset of selected AEs is defined as: (First date of selected AEs occurred – date of first dose + 1)/30.4375.

Time to Resolution of Selected AEs:

Time (Months) to resolution for an event is defined as time from occurrence of the AE to its complete resolution, which is defined as the record of the event being resolved (as recorded on “Adverse Event” eCRF page, Outcome of AE= (Recovered/Resolved, Recovered/Resolved with sequelae)). For subjects with multiple episodes of the same event (LLT) overlapped adjacent to each other, the onset date of first occurrence of the event and the latest end date of the events are used. For example, if a subject experiencing the event onset on Day 1 and resolution on Day 7, and another event of the same LLT onset on Day 6 and resolution on Day 9, then the event assumed to have onset on Day 1 and achieving resolution on Day 9. The time to resolution will be calculated to equal 9 days for analysis. In case the AE resulted in death, the time is censored at the date of death. If the outcome of the AE is not death, the time is censored at the date of death, lost to follow-up or date of last dose of study treatment + 90 days or date of cut-off for analysis, whichever is earliest.

Immune-Related Adverse Event (irAE) Assessed by Investigator: immune-related adverse events (as recorded on “Adverse Event” eCRF page, is this an immune-related adverse event= “Yes”). The systemic Corticosteroids , immunosuppressive treatment other than Corti costeroids, and hormone replacement therapy used for irAEs will be identified as for selected AEs for SHR-1210.

Duration of systemically administered Corticosteroids:

For subjects treated with high dose (⩾ 40mg/day) of systemically administered Corticosteroids, duration (total days) of systemically administered Corticosteroids is defined as the time from first dose to last dose without overlaps. For example, if a subject treated with high dose of systemically administered Corticosteroids, first dose of a systemically administered Corticosteroids on Day 1 and last dose on Day 7, and first dose of another systemically administered Corticosteroids on Day 6 and last dose on Day 9, then the duration of systemically administered Corticosteroids will be calculated as 9 days for analysis. The first dose may be less than 40mg/day.

Duration of high dose with systemically administered Corticosteroids used in subjects is defined in days as the total number of days treated with high dose without overlaps. No imputation of missing data for the duration will be conducted.

3.3.2. Laboratory Test

Clinical laboratory assessments are performed within 7 days prior to first dose (Day -7), Day 15 Cycle 1, Day 1 and 15 on Cycle 2 and 3, Day 1 on Cycle 4 onward, End-of-Treatment visit and Safety Follow-Up site visit. Clinical laboratory (e.g., hematology, blood biochemistry) values are evaluated as appropriate.

To evaluate if there are any potential risks of drug-induced liver injury (DILI) in HCC

patients, a summary and a listing for all cases meeting the potential Hy's law criteria, defined as below:

- ALT or AST ≥ 3 × ULN (normal at baseline), or ALT or AST ≥ 3 × baseline value (abnormally elevated at baseline)

  AND

- TBIL ≥ 2 × ULN (normal at baseline), or TBIL ≥ 2 × baseline value (abnormally elevated at baseline)

  AND

- ALP < 2 × ULN (normal at baseline) or ALP < 2 × baseline (abnormally elevated at baseline)

Concurrent measurements are those occurring within 7 days after ALT/AST abnormal elevation.

3.3.3. Vital Signs

Vital signs are collected on Day -14, Day 1 of each treatment cycle, Day 15 of each treatment cycle (for subjects in the Sorafenib group, vital signs are not collected on Day 15 of each treatment cycle from Cycle 4 onwards), End-of-Treatment visit and Safety Follow-Up site visit for height (cm), temperature (Celsius), systolic blood pressure (mmHg), diastolic blood pressure (mmHg), pulse rate (bpm) and respiratory rate (resp/min).

Table 6. Criteria for Clinically Significant Abnormal Vital Signs

Vital Sign	Criteria for Abnormal Low		Criteria for Abnormal High	
Parameters	**Value**	**Change from Baseline**	**Value**	**Change from Baseline**
Body Temperature	≤ 35 ˚C	≤ -1.1 ˚C	≥ 38.3 ˚C	≥ 1.1 ˚C
Pulse Rate	< 40 bpm	--	> 120 bpm	--
Respiratory Rate	< 12 resp/min	--	> 24 resp/min	--
Systolic Blood Pressure	< 90 mmHg,	≤ -30 mmHg	≥ 160mmHG	≥ 30 mmHg
Diastolic Blood Pressure	< 50 mmHg,	≤ -20 mmHg	≥ 100mmHg	≥ 20 mmHg

3.3.4. 12-lead ECG

ECG measurements are collected on Day -14, at 2.5 hours (±1 hour) after rivoceranib administration on Day 1 and Day 15 of Cycle 1 and Day 1 of subsequent treatment cycles, End-of-Treatment visit and Safety Follow-Up site visit for heart rate (bpm), PR interval (msec), QT interval (msec) and QTcF interval.

ECG accompanied blood sampling: blood are collected within 15 minutes after the ECG examination on the first day from C1 to C4.

Table 7. Criteria for Markedly Abnormal 12-lead ECG

Parameters	Measurement	Criteria for Markedly Abnormal Value
PR	Absolute value	Max ≥ 300 msec
	Change from Baseline	Baseline >200 msec and change from Baseline ≥ 25%, or Baseline ≤ 200 msec and change from Baseline ≥ 50%
QTcF	Absolute value	450 ≤ max.<480 msec
		480 ≤ max.<500 msec
		max. ≥ 500 msec
	Change from Baseline	30 ≤ max. <60 msec
		max. ≥ 60 msec

3.3.5. Physical Examination

Physical examination are performed within 14 days prior to first dose during the treatment period (experimental arm: prior to administration on Day 1 and 15 of each cycle of therapy; control arm: prior to administration on Day 1 and 15 of each cycle in Cycle 1-3, prior to administration only on Day 1 from Cycle 4), at end-of-treatment visit and safety follow-up.

Physical examination includes height (first collection only), weight, head and face, skin system, lymph nodes, eye, otolaryngology, mouth, respiratory system, cardiovascular system, abdomen, genitourinary system, musculoskeletal, neurological and mental status. If necessary, OCT (Optical Coherence Tomography) examinations are performed.

3.3.6. ECOG-PS score

ECOG-PS score are performed within 14 days prior to the first dose during the screening period, prior to administration on Day 1 of the subsequent cycles of therapy, at end-of-treatment visit and safety follow-up.

3.3.7. Child-Pugh

Child-Pugh class are performed within 7 days prior to the first dose during the screening period, on Day 1 of each cycle in Cycles ≥ 2 during the treatment period, at end-of-treatment visit and safety follow-up.

3.3.8. Echocardiography

Echocardiography must include LVEF and can be scheduled at the investigator's discretion during the study.

3.3.9. Treatment Exposure

Exposure for SHR-1210 will be summarized by cycle and overall. The start date and end date of actual cycle for calculating cycle duration are defined as following.

Start date for cycle X

- the earliest start date of dosing in the Cycle X, if the subject received study treatment in that cycle
- the first day of assessments in the Cycle X, if the subject did not receive study treatment in that cycle. Use start date in the exposure page if available; if start date is not available then use date of collection of vital signs in Cycle X day 1 visit.

End date for cycle X:

- all cycles X except for the last cycle:

  end date of actual cycle = start date of actual cycle (X+1) – 1 day;

- the last cycle:

  end date of actual cycle = start date of actual cycle + 28 – 1 day

The exposure data for Rivoceranib/Sorafenib are not collected by cycle. Overall exposure for Rivoceranib/Sorafenib will be summarized.

Table 8. Derivation of Exposure

	SHR-1210	**Rivoceranib**	**Sorafenib**
Drug administered	200 mg, via intravenous infusion, once every two weeks (Q2W), 28 days per cycle	250 mg, p.o., once per day (QD), 28 days per cycle	400 mg, p.o., twice per day (BID), 28 days per cycle
Duration of exposure (weeks)	(last dose date of SHR-1210 – first dose date of SHR-1210 + 14)/7	(last dose date – first dose date + 1)/7	
Duration of exposure (months)	(last dose date of SHR-1210 – first dose date of SHR-1210 + 14)/30.4375	(last dose date – first dose date + 1)/30.4375	
Intended duration of treatment (weeks)	(end date of SHR-1210 – first date of SHR-1210 +1)/7 where end date = start date of last cycle with non-zero dose of SHR-1210 + 28 – 1	(last dose date – first dose date + 1)/7	
Intended duration of treatment (months)	(end date of SHR-1210 – first date of SHR-1210 +1)/30.4375 where end date = start date of last cycle with non-zero dose of SHR-1210 + 28 – 1	(last dose date – first dose date + 1)/30.4375	
Cycle duration (weeks)	(end date of actual cycle – start date of actual cycle + 1)/7	-	
By cycle actual DI (mg/4-week cycle)	cumulative dose in a cycle (mg)/ [cycle duration (weeks)/4]	-	
Overall actual DI (mg/4-week cycle)	[overall cumulative dose (mg)] / [intended duration of treatment (weeks)/4]		
Intended DI (mg/4-week cycle)	200*2=400	250*28=7000	400*2*28=22400
By cycle RDI (%)	100*[by cycle actual DI] / Intended DI	-	
Overall RDI (%)	100*[overall actual DI] / Intended DI		

DI= Actual Dose Intensity; RDI= Relative Dose Intensity

3.4. Pharmacokinetic and Immunogenicity

- Serum concentration of SHR-1210 and plasma concentration of rivoceranib.
- The number and proportion of subjects with anti-SHR-1210 antibody (ADA) positive or neutralizing antibody (Nab) positive.

Efficacy endpoints (ORR and PFS assessed by BIRC, OS) and safety endpoint (e.g., drug exposure, TEAEs, treatment related AEs, TEAEs with CTCAE grade ≥3, treatment related AE with CTCAE grade ≥3, SAE, treatment related SAE, TEAEs leading to discontinuation form any study drug, TEAEs leading to discontinuation form SHR-1210, AE Leading to Dose Reduction/Interruption of any study drug) will be summarized according to the ADA status of SHR-1210 in the experimental arm.

3.5. Exploratory Endpoints

3.5.1. Patient Reported Outcome (PRO) Endpoints

EORTC QLQ-C30 is a core scale in the quality of life measurement questionnaire (QLQ) system for all cancer subjects that is systematically developed by European Organization for Research and Treatment of Cancer (EORTC), and used for measurement of quality of life (QoL) in all the cancer subjects (for the common part). And the specific scale for different cancers is formed, HCC module (EORTC QLQ-HCC18) is used in this study.

EORTC QLQ-C30 (version 3.0), EORTC QLQ-HCC18 and EQ-5D-5L questionnaires will be completed by subjects prior to administration of study drug and any other study evaluation at clinical centers on Day 1 of each cycle, End-of-Treatment visit and Safety Follow-Up visit whilst on site visit. For subjects who completed self-evaluation within 7 days of their End-of-Treatment visit shall not be required to complete them again for the visit.

3.5.1.1. EORTC QLQ-C30

The EORTC QLQ-C30 questionnaire comprised of a global health status scale for health status/QoL during the past week on a scale of 1=Very poor to 7=Excellent, five functional scales (physical, role, cognitive, emotional, social on a 4-point scale of 1=Not at all, 2=A little, 3=Quite a bit and 4=Very much), three symptom scales (fatigue, pain, nausea and vomiting) and a number of single items assessing symptoms commonly reported by cancer subjects (e.g., dyspnoea, insomnia, appetite loss, constipation, diarrhea) and perceived financial impact of the disease, all of which on a 4-point scale. All questions, except for the questions of physical functioning, are related to experiences during the past week. The 30 items are grouped by global health status / QoL scale, functional scales, symptom scales and six single items. A linearly transformed score between 0-100 for each of the scales or item will be calculated following the EORTC scoring manual [Fayers et al 1999]. A high scale score represents a higher response level. Thus, a high score for the functional scale represents a high / healthy level of functioning, a high score for the global health status / QoL represents a high QoL, but a high score for the symptom scale / item represents a high level of symptomatology / problems.

Baseline will be defined as the last non-missing assessment prior to randomization for symptoms and summaries.

The HRQoL/global health status will be assessed using the EORTC-QLQ-C30 global QoL scale which includes 2 items from the QLQ-C30: "How would you rate your overall health during the past week? (Item 29) and "How would you rate your overall QoL during the past week? (Item 30).

Definition of Clinically Meaningful Changes

Changes in score compared with baseline will be evaluated. A minimum clinically meaningful changes is defined as a change in the score from baseline of ≥10 for the EORTC QLQ-C30 (Osoba et al 1998). For example, a clinically meaningful deterioration is defined as a decrease in the score from baseline of ≥10.

In practical terms, if items $I_1, I_2, \ldots I_n$ are included in a scale, the procedure is as follows:

Raw Score

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

Linear Transformation

Apply the linear transformation to 0-100 to obtain the score S,

Functional scales： S = (1-(RS- 1) / range) * 100

Symptom scales / items: S = ((RS-1) / range) * 100

Global health status / QoL: S = ((RS-1) / range) * 100

Where *RS* is the mean of the component items. *n* is the number of items have been answered in a scale, *I* is the value of a corresponding item. *Range* is the difference between the maximum possible value of RS and minimum possible value. Most items take values from 1 to 4, giving *range* = 3.

Examples:

Emotional functioning $RS = (Q_{21} + Q_{22} + Q_{23} + Q_{24}) / 4$

EF S = (1-(RS- 1) / 3) * 100

Fatigue RS = (Q10 + Q12 + Q18) / 3

FA S = ((RS-1) / 3) * 100

Table 9：Scoring the EORTC QLQ-C30 version 3.0

Scale	Number of items	Item range	Version 3.0 Item numbers	Function scales

Global health status/QoL					
Global health status/QoL (revised) †	QL2	2	6	29,30	
Functional scales					
Physical functioning (revised)†	PF2	5	3	1-5	F
Role functioning (revised)†	RF2	2	3	6-7	F
Emotional functioning	EF	4	3	21-24	F
Cognitive functioning	CF	2	3	20,25	F
Social functioning	SF	2	3	26,27	F
Symptom scales / items					
Fatigue	FA	3	3	10,12,18	
Nausea and vomiting	NV	2	3	14,15	
Pain	PA	2	3	9,19	
Single item					
Dyspnoea	DY	1	3	8	
Insomnia	SL	1	3	11	
Appetite loss	AP	1	3	13	
Constipation	CO	1	3	16	
Diarrhoea	DI	1	3	17	
Financial difficulties	FI	1	3	28	

For subjects who failed to answer a few questions, and if at least half of the items from the scale have been answered, then use all the items that were completed and apply the standard approach for calculating the scale scores (ignore any items with missing values when making the calculations). Otherwise, set the scale score to missing. For the single-item measures, set score to missing.

Time to Deterioration (TTD)

For HRQoL/function, TTD will be defined as the time from the date of randomization to the date of first deterioration, as determined by following subscales of European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire-Core 30 (QLQ-C30) (i.e., a decreased in the function scales or the global health status/HR QoL by $\geq$ 10 from baseline maintained for two consecutive time points, or one time-point followed by death [from any cause] within 4 weeks), regardless of whether the subject withdraws from study treatment or receives another anti-cancer therapy prior to HR QoL/function deterioration, will be calculated for the following three scales:

- HRQoL/Global health status;

- Physical Functioning;
- Role Functioning;

For the selected symptoms scales/items (more HCC-related), TTD will be defined as the time from the date of randomization to the date of first deterioration (i.e., an increase in the score by $\geq$ 10 from baseline maintained for two consecutive time points, or one time-point followed by death [from any cause] within 4 weeks), regardless of whether the subject withdraws from study treatment or receives another anti-cancer therapy prior to deterioration, will be calculated for the following scales.

- Selected symptom scales / items (Fatigue, Nausea and vomiting, Pain, Appetite loss, Diarrhoea)

A change of 10 to 20 points in score represents a “very much” change.

Subjects whose HRQoL/function/symptom (as measured by EORTC-QLQ-C30) have not shown a confirmed deterioration and who are alive at the time of the analysis will be censored at the time of the last PRO assessment where the HRQoL/function/symptom documenting no deterioration. Also, if HRQoL/function/symptom deteriorates after 2 or more missed PRO assessment visits, the subject will be censored at the time of the last PRO assessment where the HRQoL/function/symptom documenting no deterioration (prior to the two missed assessment visits). Given the scheduled visit assessment scheme (on day 1 of each cycle, 28 days per cycle),two missing visits will equate to 8 weeks + 3 days. If subjects have no evaluable visits or do not have baseline data, they will be censored at day 1.

3.5.1.2. EORTC QLQ-HCC18

The EORTC QLQ-HCC18 questionnaire comprised of 18 items (Q_{31}-Q_{48}) assessing symptoms or problems during the past week on a 4-points scale of 1=Not at all, 2=A little, 3=Quite a bit and 4=Very much. Q_{48} is based on experiences during the past four weeks. The 18 items are grouped by six domains (fatigue, body image, jaundice, nutrition, pain and fever) and single items for abdominal swelling and sex life (Table 10). A linearly transformed score between 0-100 for each of the scales or item will be calculated. A high scale score represents a higher response level.

The calculation of scores for each scales and definition of Time to deterioration (TTD) is the same as EORTC QLQ-C30.

Table 10 Scoring the QLQ-HCC18

	Scale	Number of items	Item range	QLQ-HCC 18 Item numbers
Symptom scales				
Fatigue	FATI	3	3	45-47
Body image	BI	2	3	33, 35
Jaundice	JAUN	2	3	36, 37

Nutrition	NUTR	5	3	31, 32, 42-44
Pain	PAIN	2	3	38, 39
Fever	FEV	2	3	40, 41
Single items				
Abdominal swelling	AB	1	3	34
Sex life	SX	1	3	48

3.5.1.3. EQ-5D-5L

The EQ-5D-5L comprises two components: the utility index (UI) and the EQ-VAS. The UI is calculated from patient scoring of five dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression). For each dimension, subjects are asked to mark between 1: 'no problems' to 5: "unable to/extreme problems". The responses are combined to produce a five-digit number describing the participant's health status (ranging from 11111 for best possible health to 55555 for worst possible health). This is converted to a UI score based on an EQ-5D-5L value set elicited from general population samples (the base case will be the Chinese valuation set). A subject with health state 12345 indicates no problems with mobility, slight problems with self-care, moderate problems with doing usual activities, severe pain or discomfort and extreme anxiety or depression. For the dimensions without assessment, no UI score will be calculated.

For the EQ-VAS, subjects are asked to record their self-rated health on a vertical Visual Analogue Scale (VAS) with the end points 'The worst health you can imagine' and 'The best health you can imagine' at the bottom ('0') and top of the scale ('100'), respectively. Hence, an improvement in HRQoL is associated with an increase in UI and EQ-VAS.

3.5.2. imRECIST Endpoints Evaluated by Investigator

The imRECIST criteria will be used in this study for exploratory assessments of antitumor activity.

Immune-modified PD (imPD) needs to be confirmed by a second, consecutive assessment at least 4 weeks apart. imPD is also considered to be confirmed if the subject

- dies after the initial observation of imPD, or
- imPD followed by no additional assessments

Immune-modified Best Overall Response (imBOR) will be derived based on reported lesion responses at different evaluation time points from the date of randomization until immune-modified disease progression per imRECIST, according to the following rules. Best response may occur after any number of PD assessments. Only tumor assessment performed on or before the start of any further anti-cancer therapies will be considered in the assessment of imBOR.

- imCR: at least two determinations of imCR at least 4 weeks apart and before imPD defined as following. It is reasonable to consider a subject with time point responses of imCR-imNE-imCR, imCR-imNE-imNE-imCR as a confirmed complete response.

- imPR: at least two determinations of imPR or better at least 4 weeks apart and before imPD defined as following (and not qualifying for a imCR). It is reasonable to consider a subject with one imNE or two imNEs, or one imSD occurred between the two response as a confirmed partial response.

- imSD: at least one imSD assessment or better ≥ 8 weeks ±1 week after the date of randomization and before imPD defined as following (and not qualifying for imCR or imPR)

- imPD: at least two consecutive assessments by investigator. Confirmations of imPD at least 4 weeks apart. If the time point response at the subsequent scan of any imPD is imRECIST SD/PR/CR, an imRECIST PD is not considered an imPD.

- imNE: all other cases

Immune-Modified ORR (imORR) Evaluated by Investigator is defined as the percentage of subjects with imCR and imPR evaluated by investigator based on imRECIST.

Immune-Modified DCR (imDCR) Evaluated by Investigator is defined as the percentage of subjects with imCR, imPR or imSD ≥ 8 weeks ±1 week evaluated by investigator based on imRECIST.

Immune-Modified PFS (imPFS) Evaluated by Investigator is defined as the time from the date of randomization to the first occurrence of imPD or death, whichever comes first, evaluated by investigator. Radiological progression based on imRECIST not considered a PFS event if new assessment evaluated as a response or stable disease after ≥ 4 weeks; imTTP (months) = [date of event or censoring – date of randomization+1]/30.4375.

Immune-Modified TTP (imTTP) Evaluated by Investigator is defined as the time from the date of randomization to radiological progression of disease evaluated by investigator based on imRECIST. Radiological progression based on imRECIST not considered a TTP event if new assessment evaluated as a response or stable after ≥ 4 weeks; imTTP (months) = [date of event or censoring – date of randomization+1]/30.4375.

Immune-Modified DoR (imDoR) Evaluated by Investigator is defined as time from the date of first record of objective response (CR or PR) to the first occurrence of radiological progression or death, whichever comes first, evaluated by investigator based on imRECIST. Radiological progression based on imRECIST not considered a PD event of DoR, if new assessment evaluated as a response or stable after ≥ 4 weeks; imDoR (months) = [date of event or censoring – first date of mCR/imPR+1]/30.4375.

Immune-Modified TTR (imTTR) Evaluated by Investigator is defined as time from the date of randomization to the date of first record of objective response (imCR or imPR) evaluated by investigator based on imRECIST. imTTR (months) = [first date of imCR/imPR – date of randomization +1]/30.4375.

Table 11. Censoring Rules for imPFS Analyses

Scenario	Date of censoring
No baseline or no adequate post-baseline assessment	Censored at date of randomization
(imPD not confirmed or no imPD), no death and no new anti-cancer treatment initiated	Censored at last adequate tumor assessment documenting no imPD
(imPD not confirmed or no imPD), no death and new anti-cancer treatment initiated	Censored at last adequate tumor assessment documenting no imPD before new anti-cancer treatment is given
imPD (subsequently confirmed) or death after 2 or more missing or inadequate post-baseline tumor assessments	Censored at last adequate tumor assessment documenting no imPD prior to the ≥ 2 missed or inadequate post-baseline tumor assessment

Table 12. Censoring Rules for imTTP Analyses

Scenario	Date of censoring
No baseline or no adequate post-baseline assessment	Censored at date of randomization
(imPD not confirmed or no imPD), and no new anti-cancer treatment initiated	Censored at last adequate tumor assessment documenting no imPD
(imPD not confirmed or no imPD), and new anti-cancer treatment initiated	Censored at last adequate tumor assessment documenting no imPD before new anti-cancer treatment is given
(imPD not confirmed or no imPD) but death before 2 missing or inadequate post-baseline tumor assessments and with no new anti-cancer treatment	Censored at last adequate tumor assessment documenting no imPD
imPD not confirmed or no imPD /death after 2 or more missing or inadequate post-baseline tumor assessments	Censored at last adequate tumor assessment prior to the ≥ 2 missed or inadequate post-baseline tumor assessment

3.5.3. Biomarker Endpoints

- The correlation of the expression level by combined positive score (CPS, ≥1 versus <1; CPS≥5 versus <5) and tumor proportion score (TPS, ≥1% versus <1%; TPS ≥5% versus <5%) in tumor tissue with the efficacy of SHR-1210 combined with rivoceranib mesylate (ORR, PFS, OS).

4. STATISTICAL ANALYSIS METHODS

4.1. General Methods

All statistical analyses will be conducted using SAS Version 9.4 or later on the Microsoft® Windows Operating System.

All tabulations of analysis results will include summaries for the following two treatment groups: SHR-1210 IV 200 mg Q2W + Rivoceranib vs. Sorafenib, which corresponds to the randomized treatment group as following:

- Experimental group= SHR-1210 plus Rivoceranib;
- Control group=Sorafenib.

The below mentioned general principles will be followed throughout the study:

- Summary statistics for continuous variables will include the N, mean, standard deviation (SD), median, range (minimum-maximum), and 25th and 75th quartiles (Q1 and Q3).
- Categorical variables will be presented as frequency counts (N) and percentages (%) to one decimal place.
- Time-to-event variables will be summarized by Kaplan-Meier (K–M) median and 95% CI base on the Brookmeyer-Crowley method with log-log transformation; standard error will be calculated using the Greenwood formula.
- Means and medians will be presented to 1 more decimal place than the recorded data. Standard deviations and standard errors will be presented to 2 more decimal places than the recorded data.
- Minimum and maximum values will be presented using the same number of decimal places as the recorded data.
- Point estimate that may not be reached will be presented as NR=Not Reached.

Data summaries will be accompanied by individual subject data listings sorted by country/territory, study center and subject identifier. Data available from the electronic case report forms (eCRFs) will be listed. The actual day relative to start of treatment will be determined and included in listings, where appropriate. In general, study days prior to and relative to a defined time origin will be calculated as date of the time origin – date of onset and date of onset – date of the time origin + 1, respectively, or indicated otherwise. Time in months and weeks will be calculated as study days/30.4375 and study days/7, respectively. Change from Baseline is defined as post-baseline minus Baseline and Percent Change from Baseline is defined as 100*Change from Baseline/Baseline. Data in International System of Units (SI) will be provided for reporting analysis data.

4.2. Definition of Study Window and Baseline

Baseline is defined as the last non-missing value prior dosing on Day 1, Cycle 1. This includes multiple pre-dose assessments for ECOG-PS score, physical examination, vital signs, 12-lead ECG and laboratory. For repeated measurements of the same day, mean value will be used as baseline for continuous variables and "least" severity grade will be used as baseline for categorical variables.

Quality of life (QoL) questionnaires data collected on Day 1 for every treatment cycle, end-of-treatment visit and Safety Follow-Up site visit (SFU) will be mapped to analysis visit according to visit window for Baseline, Week 5 and every 4 weeks thereafter (e.g., Week 9, 13, etc.) as follows:

Table 13 Visit Window for Quality of Life Data

Visit Window	Target Day (Visit window from date of first dose)
Baseline	Study Day<=1 (Last non-missing value prior dosing on Day 1, Cycle 1)
Week 5	Study Day=29 (15 to 42)
Week 9	Study Day=57 (43 to 70)
Week 13	Study Day=85(71 to 98)
Week 17	Study Day=113 (99 to 126)
Week 21	Study Day=141 (127 to 154)
Week 25	Study Day=169 (155 to 182)
Week 29	Study Day=197 (183 to 210)
Week 33	Study Day=225 (211 to 238)
Week 37	Study Day=253 (239 to 266)
Week 41	Study Day=281 (267 to 294)
Week 45	Study Day=309 (295 to 322)
Week 49	Study Day=337 (323 to 350)
Week 53	Study Day=365 (351 to 378)
Etc.	Etc.

For windows containing multiple assessments, assessment closest to Target Day will be used. For multiple assessments occurred on the same date, the last assessment will be identified and used. The time origin for the study day is the date of first dose taken.

4.3. Handling of Missing Data

Except for the special cases, the following imputation rules will be applied for missing/partial date of events:

4.3.1. Cancer Diagnosis

Initial diagnosis date will be imputed as follows:

- If year is missing or date is completely missing, no imputation will be performed.

- If only day is missing, it will be imputed to the 1st day of the month.
- If both day and month are missing and the year is prior to the year of the first study treatment, the month and day will be imputed as January 1st.
- If both day and month are missing and the year is the same as the year of informed consent, the month and day will be imputed as January 1st.

4.3.2. Adverse Events

- If start date of an event is completely missing, it will be imputed by the start date of study treatment.
- If only the day of the AE start date is missing, but the year and month are the same as the year and month of the study drug treatment, then the AE start date will be replaced by the start of the study drug treatment.
- If both day and month of the AE start date are missing but the start year is equal to the start of study treatment, then the start date will be replaced by the start date of study treatment.
- Otherwise, the missing onset day or missing onset month will be replaced by 1.
- If the day of end date is missing, then the AE end date will be replaced by the last day of the month. If both day and month of the AE end date are missing then it will be replaced by December 31 of the year. However, any resultant date after subject' death, the date of death will be used to impute the end date. Otherwise, the end date will not be imputed.

4.3.3. Concomitant Medications

- If the start date is missing completely, then the start date will be replaced by the start of study treatment.
- If the day of start date is missing, but the month and year are equal to the start of study treatment, then the date will be replaced by the start of study treatment.
- If the day and month of the start date are missing but the start year is equal to the start of study treatment, then the medication date will be replaced by the start of study treatment. Otherwise, the missing day or month will be replaced by 1.
- If the day of end date is missing, then the end date will be replaced by the last day of the month. If both day and month of the AE end date are missing then it will be replaced by December 31 of the year. if resulting in a date later than the date of subject' death, the date of death will be used to impute the end date. Otherwise, the end date will not be imputed.

4.3.4. Study Treatment Date

If the first date of dose is completely or partially missing, no imputation will be performed. If the last date of study treatment is completely or partially missing, the

imputations will be performed as follows (the last date of study treatment will be used to impute for the start date of new anti-cancer therapy):

- If the last date of study treatment is completely missing and there is no end date of treatment on end of treatment (EOT) eCRF page and no death date, the subject should be considered to be ongoing and use the cut-off date for the analysis as the last date of study treatment. For active group, end of treatment is all study drugs discontinued.

- If the last date of study treatment is completely or partially missing, and an End date of Treatment on eCRF page or a death date is available before the cut-off date of analysis, then last date of study treatment is imputed as following:

    - IF only *Year* is available and *Year* < Year of min (EOT date, death date), THEN imputed date= 31DECYYYY;

    - IF both *Year* and *Month* are available and *Year* = Year of min (EOT date, death date) and Month < the month of min (EOT date, death date), THEN imputed date= Last day of the month;

4.3.5. Start Date of New Anti-Cancer Therapy

Start date of new anti-cancer therapy (drug therapy (except traditional Chinese medicine), local regional therapy, surgery, radiotherapy) is used for censoring in efficacy analyses. The start date of new anti-cancer therapy is the earliest start date of anti-cancer therapy recorded in the “Follow-up” eCRF pages (including concomitant anti-cancer therapy etc.,) for drug therapy (except traditional Chinese medicine), local regional therapy, surgery, radiotherapy after randomization.

The PD date, last date of study treatment and end date of new anti-cancer therapy will be used to impute the start date of new anti-cancer therapy if it is missing. And the definitions for the dates as following:

- PD date is PD date by investigator assessment.

- Last date of study treatment (see section 4.3.4).

- End date of new anti-cancer therapy is derived as follows:

    - If end date is completely missing, then ignore it

    - If only *year* is present for end date, then set to 31DECYYYY

    - If *month and year* are present for end date, then set to the last day of the month for MMMYYYY

The imputed start date of new anti-cancer therapy is derived as follows:

    - If Start date of new anti-cancer therapy is completely missing

start date = min [max (PD date + 1, last date of study treatment + 1), end date of new anti-cancer therapy]

- If Only *year* for start of anti-cancer therapy is present

  - IF *year* < year of min [max (PD date + 1, last date of study treatment + 1), end date of new anti-cancer therapy], THEN start date = 31DECYYYY;

  - IF *year* = year of min [max (PD date + 1, last date of study treatment + 1), end date of new anti-cancer therapy], THEN start date = min [max (PD date + 1, last date of study treatment + 1), end date of new anti-cancer therapy]

  - IF *year* > year of min [max (PD date + 1, last date of study treatment + 1), end date of new anti-cancer therapy], THEN start date = 01JANYYYY

- If *year and month* for start of anti-cancer therapy are present

  - IF *year* < year of min [max (PD date + 1, last date of study treatment + 1), end date of new anti-cancer therapy],

    Or

    IF *year* = year of min [max (PD date + 1, last date of study treatment + 1), end date of new anti-cancer therapy], and *month* < Month of min [max (PD date + 1 day, last dose of study treatment + 1 day), end date of new anti-cancer therapy]

    THEN start date = Last day of month for MMM YYYY;

  - IF *year* = year of min [max (PD date + 1, last date of study treatment + 1), end date of new anti-cancer therapy], and *month* = month of min [max (PD date + 1 day, last dose of study treatment + 1 day), end date of new anti-cancer therapy], THEN start date = min [max (PD date + 1 day, last date of study treatment + 1 day), end date of new anti-cancer therapy]);

  - IF *year* = year of min [max (PD date + 1, last date of study treatment + 1), end date of new anti-cancer therapy], and *month* > month of min [max (PD date + 1 day, last date of study treatment + 1 day), end date of new anti-cancer therapy]

    Or

    IF year > year of min [max (PD date + 1, last date of study treatment + 1), end date of new anti-cancer therapy],

    THEN start date = 01 MMM YYYY.

4.3.6. Death Date

Last survival date will be used to impute the partial or complete missing death date:

- If *year* is missing or date is *completely missing*: death date= the date of last survival +1 day
- If only *day* is missing: death date= max (the first of month and year of death, date of last survival + 1)
- If *day and month* are missing:
  - If *year* = year of the date of last survival, death date= the date of last survival +1 day
  - If *year* > the date of last survival, death date= 01JANYYYY

4.3.7. Pharmacokinetics

For PK concentration summaries and graphical displays, values below the quantitation limit (BQL) will be set to zero. If all concentrations at a given time point are BQL values, the mean will be presented as zero and both SD and CV% will be reported as NA (Not Applicable). BQL values will be presented as "BQL" in data listings.

If all concentrations at a given time point are missing, then summary statistics will be presented as NA.

If blood sample was collected at a nominal time and dose interruption of SHR-1210 occurred at the last administration before the nominal time, the concentration at the nominal time point will be excluded from the descriptive statistics summary and mean concentration-time plotting.

4.4. Analysis Populations

4.4.1. Intent-to-Treat Set (ITT):

This population will include all randomized subjects. Subjects will be included in the treatment group to which they are randomized.

4.4.2. Per-protocol Set (PPS)

This population is a subset of the ITT population consists of subjects who take at least one dose of study drug and have no important protocol deviations during the study that do have a significant impact on study results (see Section 4.5 for further detail).

4.4.3. Safety Analysis Set (SS)

This population will include all treated subjects who receive at least one dose of study treatment. Safety data will be analyzed based on the first treatment they actually received.

4.4.4. Anti-drug Antibody Analysis Set (ADAS)

This population will include all subjects who receive SHR-1210 and had baseline ADA measurement and at least one post-baseline ADA measurement.

4.4.5. PK Analysis Set (PKS)

This population will include all subjects who receive SHR-1210 combined with rivoceranib and have at least one measurable drug concentration at any time during the study.

4.4.6. PK/QTc Analysis Set (PKQTS)

This population will include all subjects who have at least 1 pair of post-dose PK concentration of rivoceranib and post-dose ΔQTcF data. The PK/QTc Set will be used for the concentration-QTc analysis.

4.4.7. Patient Reported Outcomes (PRO) Analysis Set

Patient reported outcomes analysis set will include subjects who have both a baseline PRO assessment and at least one post-baseline PRO assessment, with subjects grouped according to the treatment assigned at randomization.

The analysis sets will be descriptively summarized for each treatment group and total. Listing of subjects included/excluded from each of the analysis sets will be provided.

4.5. Protocol Deviations

All protocol deviations will be identified during study monitoring and important protocol deviations related to the study inclusion/exclusion criteria, study conduct, subject management or subject assessment will be identified, reviewed and categorized prior to database lock for analysis. Important protocol deviations are a subset of protocol deviations that may significantly impact the completeness, accuracy, and/or reliability of key study data (e.g., the primary study endpoints) or that may significantly affect a subject's rights, safety, or well-being. For example, important protocol deviations may include enrolling subjects in violation of key eligibility criteria or failing to collect data necessary to interpret primary endpoints, as this may compromise the scientific value of the trial. The following general protocol deviations will be used for study team reviews:

- Eligibility and Entry criteria deviation (e.g., subject deviated from inclusion/exclusion criteria);
- Prohibited medication deviation (e.g., subject received other investigational products for anti-cancer therapy during the treatment period of this study);
- Subjects received incorrect study treatment or dose;
- Other categorizations may also be identified during study reviews.

The following outputs will be provided:

- Summary of important protocol deviations by category;
- Summary of important deviations, which result in exclusion of subjects from the Per Protocol set (PPS) by category.

- Listing of protocol deviations with important deviations identified for exclusion of Per Protocol set (PPS) during study reviews;
- Summary of important protocol deviations related to COVID-19 epidemic measures by category;
- Summary of protocol deviations related to COVID-19 epidemic measures by category;
- Listing of protocol deviations related to COVID-19 epidemic measures with important deviations identified;
- Listing of important protocol deviations;

4.6. Subject Disposition

The number of subjects screened, randomized, treated and discontinued will be counted. The study drug and/or study discontinuation will be summarized according to the categories in the eCRF.

The end of study status (Physician Decision, Withdrawal by Subject, Lost to Follow-up, Death, Study Terminated by Sponsor and Other) will be summarized by treatment group and total.

The date of end of study is defined as the death date for subjects who died and the last contact date for subjects who were lost to follow-up.

The primary reasons for study drug discontinuation including AE (subgroup of AE (non-COVID-19), AE (COVID-19)), COVID-19 (non-AE), Disease Progression, Withdrawal by Subject, Physician Decision, Protocol Deviation, Study Terminated by Sponsor, Lost to Follow-Up, Death, Pregnancy, and Other will be summarized for each drug component by treatment group.

Duration of subject follow-up time (months) for each treatment group and total will be summarized using descriptive statistics, including 25th and 75th quartiles (Q1 and Q3).

The results of the randomization according to IRT will be summarized as below:

- Number and percentage of randomized subjects by Country and center
- Number and percentage of randomized subjects by randomized strata from IRT and eCRF, separately

A subject listing of stratum by IRT and stratum by eCRF will be provided.

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

A subject listing for all treated subjects will be provided showing the subject's randomization date, first and last dosing date, date investigator decided to stop the treatment and reason for discontinuation.

4.7. Demographics and Baseline Characteristics

Demographic and other baseline characteristics captured during screening period will be summarized using descriptive statistics. Continuous variables include age, vital signs including weight, height, and BMI, etc. History of HCC diagnosis and characteristics will be summarized using descriptive statistics for time since initial cancer diagnosis (days), time since advanced/metastatic disease diagnosis (days). Categorical variables include, gender, ethnicity, ECOG-PS score, country, geographical region (Asia vs. non-Asia countries), race, race group (Asian vs. non-Asian [White vs. African American vs. others] as listed in the eCRF), hepatitis virus, age group (< 65 vs. ⩾65 years), weight group (<60 vs. ⩾60 kg), status of alcohol consumption (absence vs. presence), BMI group (<30 vs. ⩾30 kg/m^2), BCLC stage at Screening, macrovascular invasion status (yes vs. no), extrahepatic metastasis status (yes vs. no), macrovascular invasion and/or extrahepatic metastasis (yes vs. no) status, root cause of HCC (HBV, HCV and Non-viral, Non-viral included Alcohol Cirrhosis, Genetic, Metabolic Disorders (such as NASH / AFLD) and others), Child-Pugh classification, AFP group (<400 ng/mL vs. ⩾ 400 ng/mL) as listed in the eCRF. Country and region will be derived from location of subject via Subject ID for analysis. PD-L1 category will include PD-L1 categories (TPS <1% vs. TPS >=1%, TPS<5% vs. TPS >=5%, and CPS <1 vs. CPS >=1, CPS <5 vs. CPS >=5). Height and weight values will be presented in metric units (cm and kg, respectively), and BMI is calculated as [weight (kg) / (height (m))^2] using the weight and height collected during the vital signs at Screening period of the study.

Tabulations by study center, country or subgroup may be repeated, as indicated, including additional tables maybe produced to support local registrations such as China, USA and Europe. Countries/regions in Asia included in the study are Mainland China, Hong Kong, Taiwan and South Korea.

4.8. Medical History

Medical history will be coded using the MedDRA (v 24.0 or higher) for lower level term (LLT), preferred term (PT) and primary system organ class (SOC). The medical history will be summarized by SOC and PT, and listed.

4.9. Prior and Subsequent Anti-Cancer Therapy

Prior anti-cancer therapy will be provided in a data listing with data retrieved from 'Prior Anti-Cancer Drug Therapy', 'Prior Local Regional Therapy', 'Prior Anti-Cancer Related Radiotherapy' and 'Prior Cancer-Related Surgery/Procedure' eCRF pages.
Subsequent anti-cancer therapy will be provided in a data listing with data retrieved from 'Follow-up Anti-Cancer Drug Therapy', 'Follow-up Local Regional Therapy', 'Follow-up Anti-Cancer Related Radiotherapy' and 'Follow-up Cancer-Related Surgery/Procedure' eCRF pages after the last dose date.

Subjects who received prior anti-cancer drug therapy, prior local regional therapy, prior cancer-related surgery/procedure, prior anti-cancer related radiotherapy will be summarized using descriptive statistics (n and %) and listed. Similar analysis will also be performed for subsequent anti-cancer treatment.

4.10. Prior and Concomitant Therapy

Prior and concomitant therapies collected from Screening to end of safety follow-up will be coded using the WHO Drug Dictionary. Prior and concomitant medications will be summarized and listed by drug and drug class. Prior medications will be defined as medications taken within 28 days of first dose of study drug that were stopped prior to study drug administration. Concomitant medications will be defined as medications that 1) started before the first dose of study drug and continued during the on-treatment period, or 2) started during the on-treatment period. Prior and concomitant medications will be listed.

A separate listing for medications related to treatment of COVID-19 infection will also be provided.

4.11. Efficacy Analyses

The two primary efficacy endpoints for the study are OS and PFS (assessed by BIRC base on RECIST v1.1). For the PFS, there is one planned analysis, which is expected to occur when approximately 332 PFS events are observed (approximately 19 months after first subject enrolled). The OS will be analyzed in the following time points: when PFS is analyzed (approximately 19 months after first subject enrolled); when approximately 251 OS events (70% are observed at approximately 23 months after first subject enrolled) and when 359 OS events are observed (approximately 36 months after first subject enrolled), respectively.

The primary endpoints of PFS assessed by BIRC and OS, key secondary endpoint ORR will be analyzed based on ITT and PPS. The primary analysis is based on ITT. The sensitivity analysis is based on PPS. Other secondary and exploratory efficacy endpoints will be analyzed based on ITT.

Detail of timing, reasons, number of events and multiplicity adjustment methods for the PFS, OS and ORR endpoint analyses can be found in the sections for interim analysis and multiplicity.

Listing of BOR, DoR, PFS assessed by BIRC according to RECIST 1.1 and OS will be provided.

4.11.1. Primary Efficacy Endpoints Analyses

4.11.1.1. Progression-Free Survival by BIRC assessment

Primary Analysis

The definition of PFS endpoint by BIRC based on RECIST v1.1 is in section 3.1.2.

The PFS by BIRC assessment based on ITT will be evaluated by comparing SHR-1210 plus rivoceranib to sorafenib on PFS events in the intention-to-treat (ITT) population using a stratified log-rank test (based on randomization stratification factors). The hazard ratio (HR=SHR-1210 plus rivoceranib / sorafenib) and the corresponding 95% confidence interval (CI) will be estimated using a stratified Cox proportional hazard model with treatment and randomization stratification factors

included in the model. The stratification factors are macrovascular invasion and/or extrahepatic metastasis (presence vs. absence), geographical region (Asia vs. countries outside of Asia) and Baseline AFP (AFP < 400 ng/mL vs. AFP ⩾400 ng/mL). The values of the factors will be obtained from the interactive response technology [IRT] system.

The PFS for each treatment group will be estimated using the Kaplan-Meier (KM) product-limit method. The 95% CIs for median PFS and other quartiles will be estimated by Brookmeyer-Crowley method for each treatment group. The PFS rates at 3, 6, 9, 12 months will be estimated by Kaplan-Meier method and their corresponding 95% CIs will be calculated using the log (-log) transformation (based on normal approximation) with back transformation to CIs on the untransformed scale. The censoring rule for the primary analysis is described in Table 2 and frequency of subjects (n, %) experiencing PFS events and censored by reason will be provided.

Superiority of SHR-1210 plus rivoceranib over sorafenib in PFS is demonstrated, if one-sided p-value obtained from the stratified log-rank test is significant at 0.005 one-sided in favoring the experimental group. If PFS is non-significant at the 0.005 one-sided but OS endpoint is significant at the overall 0.020 one-sided, then the PFS can be tested again at an updated 0.025 one-sided after alpha re-allocation. No alpha re-allocation is performed if both PFS and OS endpoints failed to show a significant result (see multiplicity section for alpha re-allocation).

Sensitivity Analyses

- To assess the impact of the randomization stratification factors, the analysis of PFS by BIRC assessment will be repeated without the stratification factors included in the analysis.

- A sensitivity analyses will be performed using the randomization strata derived from the eCRF instead of the randomization strata obtained from the interactive response technology (IRT) system.

- A sensitivity analysis will be performed by treating documented PD/death after two or more missed disease assessments as an event.

- A sensitivity analysis will be performed by treating documented PD/death occurring on or after any new anti-cancer treatment as an event.

- A sensitivity analysis of interval-censored data will be performed in case an imbalance between the treatment groups on tumor assessment schedules is observed. In this analysis, a Finkelstein's likelihood-based score test will be used to compare the two treatment groups. The interval extends from the date of the last disease assessment without documented PD to the date of documented PD or death, whichever occurs earlier. Subjects without documented PD or death will be right-censored according to the censoring rule for the primary PFS analysis.

- The proportional hazards assumption will be checked visually by plotting log (-log (PFS)) versus log(time) within each randomization stratum. The Schoenfeld residuals for the stratified Cox proportional regression model will be plotted to

investigate graphically violations of the proportional hazards assumption; a non-zero slope is evidence of departure from proportional hazards. If it suggests the proportional hazard assumption is not met, the PFS by BIRC assessment will be analyzed by a Wilcoxon test and the restricted mean survival time (RMST) analyses will be performed as follows.

- A sensitivity analysis using a Wilcoxon test will be performed for the PFS by BIRC assessment.

- A sensitivity analysis using Restricted Mean Survival time (RMST) method will also be performed for the PFS by BIRC assessment. The cut-off point (τ) will be used to evaluate the RMST, it should not exceed the minimum of the largest observed time for both experiment group and control group. To avoid arbitrary selection of cut-off point for both the two groups, two cut-off points will be used:

- τ_1 = minimum of (largest observed survival time for experimental group, largest observed survival time for control group).

- τ_2 = minimum of (largest survival event time for experimental group, largest survival event time for control group).

The Impact of COVID-19

Assessment of the impact of COVID-19 on PFS assessment will be conducted. If deemed necessary, the below analyses may be conducted.

A PFS sensitivity analysis to assess the impacted by COVID-19 pandemic measures may be performed. This analysis is the same as the primary analysis except that it censors subjects at the time of last tumor assessment prior to the earliest time of study treatment discontinuation or interruption (SHR-1210 i.v. $\geq$ 6 times or rivoceranib $\geq$ 28 days, Sorafenib $\geq$ 28 days in succession) due to COVID-19 (non-AE).

In addition, a PFS sensitivity analysis to assess the impacted by COVID-19 pandemic measures may be performed, by excluding subjects who discontinue or interrupt the study drug treatment for certain period of time (SHR-1210 i.v. $\geq$ 6 times or rivoceranib $\geq$ 28 days, Sorafenib $\geq$ 28 days in succession) due to COVID-19 (non-AE). The method of analysis is the same as the primary analysis of PFS.

4.11.1.2. Overall Survival (OS)

Primary Analysis

OS will be evaluated by comparing SHR-1210 plus rivoceranib to sorafenib on OS events in the ITT population using a stratified Log-rank test (based on randomization stratification factors). The hazard ratio (HR=SHR-1210 plus rivoceranib / sorafenib) and the corresponding 95% confidence interval (CI) will be estimated using a stratified Cox proportional hazard model with treatment and randomization stratification factors included in the model. The OS rates for each treatment group will be estimated using the Kaplan-Meier (KM) product-limit method. A 95% CI for median OS will be estimated by Brookmeyer-Crowley method for each treatment

group. The OS rates at 6, 9, 12,18 and 24 months will be estimated by Kaplan-Meier method and their corresponding 95% CI will be calculated. The stratification factors for randomization are macrovascular invasion and/or extrahepatic metastasis (presence vs. absence), geographical region (Asia vs. countries outside of Asia) and Baseline AFP (AFP < 400 ng/mL vs. AFP ⩾400 ng/mL). The values of the factors will be obtained from the interactive response technology [IRT] system.

Superiority of SHR-1210 plus rivoceranib over sorafenib in OS is demonstrated, if one-sided p-value obtained from the stratified log-rank test is significant at an overall significance level of 0.020 one-sided in favoring the experimental group. If OS is non-significant at the overall 0.020 one-sided, but PFS endpoint is significant at the 0.005 one-sided, then OS can be tested again with the efficacy boundaries revised according to an updated 0.025 significance level after alpha re-allocation and the actual number of events observed. No alpha re-allocation is performed if both PFS and OS endpoints failed to show a significant result (see multiplicity section for alpha re-allocation).

The OS analysis will be performed at the final PFS analysis, at 251 (70%) deaths if there are less than 251 deaths at the final PFS analysis at ~23 months, and when ~359 OS events are observed at ~36 months, if study continued as planned. Detail of timing, reasons, number of events and multiplicity adjustment methods for the PFS, OS and ORR endpoint analyses can be found in the sections for interim analysis and multiplicity.

Sensitivity Analyses

- To assess the impact of the randomization stratification factors, the above primary analysis of OS will be repeated without the stratification factors included in the analysis.

- A sensitivity analyses will be performed using the randomization strata derived according to eCRF instead of the randomization strata obtained from the interactive response technology (IRT) system.

- To account for possible confounding effect due to patients switching to new anti-cancer treatment (except for traditional Chinese medicine) after the last dose date, a sensitivity analysis of OS that censoring subjects at the time of initiation of the new anti-cancer therapy after the last dose will be performed.

- The assumption of proportionality will be assessed in the same way as for PFS by BIRC assessment. If proportional hazard assumption is not met, a sensitivity analysis using a Wilcoxon test will be performed for the OS.

- If proportional hazard assumption is not met, a sensitivity analysis using Restricted Mean Survival time (RMST) method will also be performed for the OS.

In addition, to further assess the impact of switching to immunotherapy, or targeted therapy (Monoclonal Antibodies Therapy, Small Molecules Target Therapy) or other treatment after the last dose of study drug on OS results, methods such as Inverse Probability of Censoring Weighting (Robins and Finkelstein 2000), and two-stage method (Latimer 2017) will be applied as appropriate. The final application of

methods will be assessed based on a blinded review of the data and the plausibility of the underlying assumptions.

The Impact of COVID-19

Assessment of the impact of COVID-19 on OS will be conducted. If deemed necessary, the below analyses may be conducted.

An OS sensitivity analysis to assess the impacted by COVID-19 pandemic measures may be performed. This analysis is the same as the primary analysis except that it censors subjects at the time of last tumor assessment prior to the earliest time of study treatment discontinuation or interruption (SHR-1210 i.v. ≥ 6 times or rivoceranib ≥ 28 days, Sorafenib ≥ 28 days in succession) due to COVID-19 (non-AE).

In addition, an OS sensitivity analysis to assess the impacted by COVID-19 pandemic measures may be performed by excluding subjects who discontinue or interrupt the study drug treatment for certain period of time (SHR-1210 i.v. ≥ 6 times or rivoceranib ≥ 28 days, Sorafenib ≥ 28 days in succession) due to COVID-19 (non-AE). The method of analysis is the same as the primary analysis of OS.

4.11.2. Secondary and Exploratory Efficacy Endpoints Analyses

4.11.2.1. Progression-Free Survival by Investigator according to RECIST and imRECIST, and by BIRC according to mRECIST Assessment

The method for the below analyses is the same as the primary endpoint PFS assessment by BIRC.

- The analysis of PFS by investigator assessment according to RECIST 1.1 will be performed with the stratification factors included in the analysis.
- The analysis of PFS by investigator assessment according to RECIST 1.1 will be performed without the stratification factors included in the analysis.
- The analysis of PFS by BIRC according to the mRECIST will be performed with the stratification factors included in the analysis.
- The analysis of PFS by BIRC according to the mRECIST will be performed without the stratification factors included in the analysis.
- The analysis of PFS by investigator assessment according to the imRECIST will be performed with the stratification factors included in the analysis.
- The analysis of PFS by investigator assessment according to the imRECIST will be performed without the stratification factors included in the analysis.

4.11.2.2. Objective Response Rate (ORR)

Confirmed BOR and ORR:

Frequency (%) of best overall response (BOR) category for each treatment group will be provided. The objective response rate (ORR) based on complete response (CR) and partial response (PR) as assessed by BIRC according to the RECIST v1.1 will be analyzed when 70% OS events are observed. Comparison between treatment groups in ORR by BIRC according to RECIST v1.1 will be performed using stratified Cochran-Mantel-Haenszel (CMH) test. Difference in proportions (%) for the ORR and its 95% CI using normal approximation will be provided. Overall response rates and their corresponding 95% exact CIs will be calculated by Clopper-Pearson method for each treatment group. Hypothesis testing of ORR will be conducted only if both PFS and OS endpoints are significant and when 70% OS events are observed.

ORR by investigator according to the RECIST v1.1, ORR by BIRC according to mRECIST and ORR by investigator according to imRECIST will be evaluated.

The analyses except for the imRECIST assessment will be repeated in the ITT subjects with measurable disease at baseline.

Unconfirmed BOR and ORR:

The unconfirmed BOR and ORR by BIRC and Investigator according to RECIST 1.1 and mRECIST will be summarized as confirmed BOR and ORR.

The analyses will be repeated in the ITT subjects with measurable disease at baseline.

Listing of confirmed ORR and unconfirmed ORR will be provided.

Change in Tumor Size

The absolute values and percentage change in target lesions (TL) from baseline will be summarized using descriptive statistics and presented at each time point by treatment group. The best change in TL tumor size will also be presented graphically using waterfall plots for each treatment group.

4.11.2.3. Duration of Response (DoR)

The DoR will be summarized descriptively using Kaplan-Meier (KM) product-limited method for median and quartile estimates with 95% CIs estimated using the Brookmeyer-Crowley method for each treatment group. DoR rates at 3, 6, 9 and 12 months obtained from the KM method will also be presented and their respective 95% CIs will be calculated. The 95% CIs will be calculated using the log (-log) transformation (based on normal approximation) with back transformation to CIs on the untransformed scale.

4.11.2.4. Time-to-Progression (TTP)

The treatment comparison for TTP will be based on a stratified Log-rank test (based on randomization stratification factors). The hazard ratio (HR=SHR-1210 plus rivoceranib / sorafenib) and the corresponding 95% confidence interval (CI) will be estimated using a stratified Cox proportional hazard model with treatment and randomization stratification factors included in the model. The TTP curve for each treatment group will be estimated using the Kaplan-Meier (KM) product-limit

method. A 95% CI for median TTP will be estimated by the Brookmeyer-Crowley method.

4.11.2.5. Disease Control Rate (DCR)

Disease control rate (DCR in %) will be analyzed using the same methods as ORR.

4.11.2.6. Time to Response (TTR)

The TTR will be summarized descriptively using n, mean, SD, median, minimum and maximum.

4.12. BIRC versus Investigator Assessment

The number of discordant assessments between BIRC assessment and investigator assessment will be provided. The number of discordant assessments is the number of PDs assessed by BIRC but not the investigator, or by investigator not the BIRC, timing is not considered.

4.13. Other Exploratory Endpoints Analyses

4.13.1. Patient Reported Outcome (PRO) Endpoints Analyses

For each treatment group and at each time point, the number and percentage of patients who complete all questions, complete at least one question and missing assessment will be summarized, completion rate (%) over the study for each questionnaire will be performed. The completion rate is defined as the proportion of total number of questions completed divided by the total number of questions attempted. Reasons for not done by analysis visit will also be provided. Reasons for missing assessment will be provided in listing.

The PRO analysis set will be used for descriptive analyses of visit summary and change from baseline and proportion analyses. The ITT set will be used for the analyses of PRO completion and time-to deterioration. Analyses by visit will be based on the Analysis Visit as defined in Section 4.2. Time-to-event analysis will be based on all assessments.

4.13.1.1.1. EORTC QLQ-C30

The linearly transformed score for each of the scales or item will be summarized descriptively, including number of missing values and standard error (SE), for observed value and change from Baseline by analysis time point (e.g., Baseline, Week 5, Week 9, etc.) by treatment group. Treatment group comparison in change from Baseline will be based on a linear Mixed Model Repeated Measures model (MMRM), baseline value as covariate, treatment group, time and treatment-by-time. Subject will be included as a random effect. An unstructured (UN) covariance structured for repeated measurements over time may be used. If the fit of the unstructured covariance structure fails to converge, the Compound Symmetry (CS) covariance structures will be used. LSMean and LSMean difference (SHR-1210 plus rivoceranib - sorafenib) with corresponding 95% confidence interval (CI) will be estimated and P

value will be reported. Plots of LSmean difference (95% CI) for change from Baseline over time will be provided. Both raw and linearly transformed scores will be listed.

For each endpoint of global health status (GHS), physical functioning, role functioning and the selected symptom scales / items (Fatigue, Nausea and vomiting, Pain, Appetite loss, Diarrhoea), the treatment comparison for Time to Deterioration (TTD) will be based on a stratified Log-rank test (based on randomization stratification factors). The hazard ratio (HR=SHR-1210 plus rivoceranib / sorafenib) and the corresponding 95% confidence interval (CI) will be estimated using a stratified Cox proportional hazard model with treatment and randomization stratification factors included in the model. Time to deterioration (TTD) curve for each treatment group will be estimated using the Kaplan-Meier (KM) product-limit method. A 95% CI for median will be estimated by the Brookmeyer-Crowley method. For this exploratory analysis, subjects with less than 10 points at Baseline will be excluded from the analysis.

4.13.1.1.2. EORTC QLQ-HCC18

The linearly transformed score for each of the scales or item will be summarized descriptively, including number of missing values and standard error (SE), for observed value and change from Baseline by analysis time point and treatment group. Treatment group comparison in change from Baseline will be based on a linear Mixed Model Repeated Measures model (MMRM), baseline value as covariate, treatment group, time and treatment-by-time. Subject will be included as a random effect. An unstructured (UN) covariance structured for repeated measurements over time may be used. If the fit of the unstructured covariance structure fails to converge, the Compound Symmetry (CS) covariance structures will be used. LSMean and LSMean difference (SHR-1210 plus rivoceranib - sorafenib) with 95% confidence interval (CI) and p-value will be calculated. Plots of LSmean difference (95% CI) for change from Baseline over time will be provided. Both raw score and linearly transformed scores will be listed.

For the selected (more HCC-related) symptom scales/items including fatigue, jaundice, pain, abdominal swelling, the treatment comparison for TTD will be based on a stratified Log-rank test (based on randomization stratification factors). The hazard ratio (HR=SHR-1210 plus rivoceranib / sorafenib) and the corresponding 95% confidence interval (CI) will be estimated using a stratified Cox proportional hazard model with treatment and randomization stratification factors included in the model. The TTD curve for each treatment group will be estimated using the Kaplan-Meier (KM) product-limit method. A 95% CI for median will be estimated by the Brookmeyer-Crowley method.

4.13.1.1.3. EQ-5D-5L

Both UI and EQ-VAS will be summarized descriptively, including number of missing values and standard error (SE), for observed value and change from Baseline by analysis time point and treatment group. Treatment group comparison in change from Baseline will be based on a linear Mixed Model Repeated Measures model (MMRM), baseline value as covariate, treatment group, time and treatment-by-time. Subject will be included as a random effect. An unstructured (UN) covariance structured for

repeated measurements over time will be used. If the fit of the unstructured covariance structure fails to converge, the Compound Symmetry (CS) covariance structures will be used. LSMean and LSMean difference (SHR-1210 plus rivoceranib - sorafenib) and the corresponding 95% confidence interval (CI) will be estimated and p-value will be calculated. Plots of LSmean difference (95% CI) for change from Baseline over time will be provided.

Listing of EQ-5D-5L will be provided.

4.13.2. Biomarker Endpoints Analysis

Based on the ITT set, the number and overall response rate by BIRC according to RECIST v1.1 for each expression level will be provided. The comparison between expression levels in ORR will be performed using the Chi-square test. Difference in proportions (%) for the ORR and its 95% CI using normal approximation will be provided. For PFS assessment by BIRC according to RECIST 1.1 and OS, the hazard ratios for the comparisons across expression levels and the corresponding 95% confidence intervals (CIs) will be estimated using a Cox proportional hazard model.

4.14. Pharmacokinetics and Immunogenicity Endpoint Analyses

4.14.1. Pharmacokinetics

The PK analysis will be conducted based on the PKS.

Serum concentrations of SHR-1210 and plasma concentrations of rivoceranib from patients in the PKS will be summarized by nominal time descriptively, including n, mean, standard deviation, median, minimum, maximum, coefficient of variation (CV%), for each drug. Possible stratifications (i.e. age, gender, body weight as categorized in subgroup section 4.18) may be applied in summary descriptive statistics.

The plots of mean (SD) concentration per drug over nominal time on linear and log scale will also be provided.

The population PK and exposure to response relationship will be analyzed and reported separately.

4.14.2. Immunogenicity Analysis of SHR-1210

The Immunogenicity analysis of ADA and Nab will be conducted based on ADAS.

4.14.2.1. Descriptive Summary of Immunogenicity

The number and percentage of all ADA and Nab classification will be summarized descriptively.

ADA classification:

(1) Non-treatment emergent ADA positive includes two situations:

a) subjects who have baseline positive samples and post-baseline negative samples;

OR

b) subjects who have both baseline and post-baseline positive samples, and the titer of the post-baseline samples is less than 4-fold of the baseline titer;

(2) Treatment-enhanced ADA positive: Subjects have both baseline and post-baseline positive samples, and the titer of the post-baseline samples is equal or more than 4-fold of the baseline titer;

(3) Treatment-induced ADA positive: subjects have baseline negative and post-baseline positive samples.

The treatment-induced ADA positive will be divided into three categories: transient ADA positive, persistent ADA positive and other ADA positive:

- Persistent ADA positive:

a) negative at baseline, there are at least two positive samples post baseline, where the first and last ADA-positive samples are separated by a period ≥16 weeks;

OR

b) negative at baseline, the last sample of post-baseline is positive or the last sample of post-baseline is negative and the second last sample is positive before an ADA-negative last sample;

- Transient ADA positive:

subjects who are negative at baseline and have only one ADA-positive sample at post-baseline before the last sample and are not persistent ADA positive;

- Other ADA positive:

subjects who are negative at baseline, have at least two ADA-positive samples post baseline, but are not persistent ADA positive, and the last sample is ADA-negative;

ADA Positive and ADA Negative：

- ADA positive (treatment emergent ADA positive): the total subjects who are treatment-induced positive and treatment-enhanced positive;
- ADA negative: subjects who are not ADA positive will be classified as ADA negative;

Nab Positive and Nab Negative：

- Nab positive: subjects who have at least one post-baseline positive Nab sample will be classified as Nab positive (Except for subjects with both baseline and post-baseline positive samples, which will be classified as Nab inconclusive);

- Nab negative: subjects who are not Nab positive and Nab inconclusive will be classified as Nab negative.

Time to and Duration of ADA and Nab response:

The ADA and Nab analyses described below will include patients with treatment-induced ADA positive or Nab positive, respectively.

Time (Weeks) to ADA response is defined as:

(Date of first ADA positive sample – date of first dosing of SHR-1210 + 1)/7.

Duration (weeks) of ADA response is defined as:

(Date of last positive ADA result – date of first positive ADA result + 1)/7.

Duration of ADA response will be censored if:

- the last ADA assessment is positive and patient is ongoing treatment with SHR-1210;
- the last ADA assessment is positive and subject discontinued treatment with SHR-1210 and the last planned ADA assessment (30 days after the last dose of SHR-1210) is after the cut-off date.

Time to ADA response will be summarized using simple descriptive statistics (mean, SD, median, min, max. quartiles).

Time to Nab response and duration of Nab response are defined similarly based on the first and the last positive Nab result.

Duration of ADA response or Nab response will be displayed graphically and analyzed using Kaplan-Meier methodology. If the number of patients with ADA response or Nab response is small, the Kaplan-Meier method may not provide reliable estimates. In this case, only descriptive statistics or listings will be provided

4.14.2.2. Demographics and Baseline Characteristics

Demographics and baseline characteristics by ADA status (ADA positive and ADA negative) will be summarized using descriptive statistics.

4.14.2.3. Impact of Immunogenicity on PK of SHR-1210

PK trough concentration of SHR-1210 at each visit stratified by ADA status (ADA positive and ADA negative) will be summarized using descriptive statistics. Also, the plot of mean concentration of SHR-1210 versus time (log and linear scale) stratified by ADA status will be drawn.

4.14.2.4. Impact of Immunogenicity on Efficacy of SHR-1210

Efficacy endpoints (ORR, PFS assessed by BIRC, OS etc.) stratified by ADA status (ADA positive and ADA negative) will be summarized using descriptive statistics, including KM estimates and plots, as appropriate.

4.14.2.5. Impact of Immunogenicity on Safety of SHR-1210

Safety endpoints (drug exposure, TEAEs, treatment related AEs, TEAEs with CTCAE grade ≥3, treatment related AE with CTCAE grade ≥3, SAE, treatment related SAE, TEAEs leading to discontinuation from any study drug, TEAEs leading to discontinuation form SHR-1210, AE leading to dose reduction/interruption of any Study drug) stratified by ADA status (ADA positive and ADA negative) will be summarized using descriptive statistics.

Selected AEs including selected AEs and grade ≥3 selected AEs (Grade 3-4, Grade 5, grade ≥3) for SHR-1210, Serious selected AEs for SHR-1210, Selected AEs for SHR-1210 leading to discontinuation from any study drug, selected AEs for SHR-1210 leading to dose reduction/interruption of any study drug will also be summarized by ADA status (ADA positive and ADA negative).

Drug exposure will also be summarized by ADA status (ADA positive and ADA negative).

ADA and Nab will be assessed as covariates in SHR-1210 population PK analysis and maybe in exposure-response analysis to see if ADA positive/ Nab positive has significant effect on PK and exposure to response relationship of SHR-1210. The population PK and exposure-response analysis will be performed and reported separately.

4.15. Safety Analyses

All the safety analyses will be based on the safety analysis set. Safety data collected during the on-treatment period will be summarized.

Number and percentage of subjects with missing visits due to COVID-19 (non-AE) will be summarized by overall and visit.

4.15.1. Extent of Exposure

Exposure will be summarized by cycle and overall for each study drug in each group. The following summaries related to study treatment will be produced for the safety set:

- Duration of exposure
- Intended duration of treatment
- By cycle actual DI
- Overall actual DI

- By cycle RDI
- Overall RDI
- Number of SHR-1210 infusions received
- Reasons for SHR-1210 with dose interruption/withdrawn
- Reasons for Rivoceranib/Sorafenib with dose interruption/withdrawn/reduced

All study drug administration and accountability data will be listed by country, study center and subject number.

Number and percentage of treatment discontinuation or interruption due to COVID-19 (non-AE) will be summarized. The number and percentage of subjects with interruption (<2, 2-6, ≥6 times of SHR-1210, duration of interruption (SHR-1210 ≥ 6 times, Rivoceranib ≥ 28 days, Sorafenib ≥ 28 days in succession) will also be summarized. The duration of interruption for each study drug will be summarized using descriptive statistics. Duration (Days) of interruption is defined as interval between the last dose date prior to an interruption and the date of first dose given after an interruption.

4.15.2. Adverse Events

4.15.2.1.1. All Adverse Events

Only TEAEs will be descriptively summarized. All AEs will be presented in subject data listings. Any AE occurring before randomized treatment (before the administration of the first dose of any drug) will be included in the AE listings and period will be referred to as “pre-treatment”.

Incidence of TEAEs will be reported as number (percentage) of subjects with TEAEs by SOC and preferred term. Subjects will be counted only once at the highest severity grade per NCI-CTCAE v 4.03 within an SOC and preferred term, even if the subject experienced more than one TEAE within a specific SOC and preferred term. The number (percentage) of subjects with TEAEs will also be summarized by relationship to study drug. Unless specified otherwise, TEAE tables will be ordered by decreasing number of subjects in SOC and PT within the SOC for experimental group, and then by alphabetically order in SOC and PT within the SOC for tied frequency. Analyses will be performed by treatment group and overall. For treatment related analyses, results by individual treatment components for the experimental group will also be presented as indicated.

In case a subject has events with non-missing and missing grades, the maximum of the non-missing grades will be displayed. No imputation of missing grades will be performed.

Subject incidence of TEAE, Treatment-emergent SAE, TEAE with CTCAE Grade ≥3, Treatment-emergent SAE with CTCAE Grade ≥3, drug related TEAE, drug related Treatment-emergent SAE, drug related TEAE with CTCAE Grade ≥3, drug related Treatment-emergent SAE with CTCAE Grade ≥3, TEAE with incidence ≥5%, drug

related TEAE with incidence ≥5%, AE leading to interruption of study treatment, AE leading to dose reduction, AE leading to discontinuation of study treatment etc. will be tabulated by treatment group.

Treatment-emergent AE summaries will present the number and percentage of subjects with at least one TEAE by treatment group. The following summaries include:

- An overview of adverse events with number and percentage of subjects, at subject level, by treatment group will be provided for the following categories:
  - All TEAEs;
  - Treatment related TEAEs with any study drug;
  - Treatment related TEAEs with SHR-1210;
  - Treatment related TEAEs with Rivoceranib / Sorafenib;
  - Grade >=3 TEAEs;
  - Grade >=3 TEAEs related to study drug;
  - Treatment-emergent SAEs;
  - Treatment-emergent SAEs related to study drug;
  - TEAEs leading to dose reduction or interruption of any study drug;
  - Treatment related TEAEs leading to dose reduction or interruption of any study drug;
  - TEAEs leading to dose reduction of Rivoceranib / Sorafenib;
  - TEAEs leading to interruption of any study drug;
  - TEAEs leading to interruption of all study drugs;
  - TEAEs leading to interruption of SHR-1210;
  - TEAEs leading to interruption of Rivoceranib;
  - TEAEs leading to interruption of Sorafenib;
  - TEAEs leading to discontinuation of any study drug;
  - Treatment related TEAEs leading to discontinuation of any study drug;
  - TEAEs leading to discontinuation of all study drugs;
  - Treatment related TEAEs leading to discontinuation of all study drugs;
  - TEAEs leading to discontinuation of SHR-1210;

- Treatment related TEAEs leading to discontinuation of SHR-1210;
- TEAEs leading to discontinuation of Rivoceranib;
- Treatment related TEAEs leading to discontinuation of Rivoceranib;
- TEAEs leading to discontinuation of Sorafenib;
- Treatment related TEAEs leading to discontinuation of Sorafenib;
- TEAEs leading to death;
- Treatment-related TEAEs leading to death
- Immun-related TEAEs assessed by investigator
- COVID-19-related TEAEs assessed by investigator

- Summary of TEAEs by SOC and PT;
- Summary of TEAEs by SOC and PT, and by maximum CTCAE grade;
- Summary of treatment-related TEAEs by SOC and PT;
- Summary of treatment-related TEAEs by SOC and PT, and by maximum CTCAE grade;
- Summary of treatment-related TEAEs with SHR-1210 by SOC and PT, and by maximum CTCAE grade;
- Summary of treatment-related TEAEs with Rivoceranib / Sorafenib by SOC and PT, and by maximum CTCAE grade;
- Summary of TEAEs with frequency $\geq$ 5% in any treatment group by SOC and PT;
- Summary of treatment-related TEAEs with frequency $\geq$ 5% in any treatment group by SOC and PT;
- Summary of Grade >=3 TEAEs by SOC and PT, and by maximum grade;
- Summary of Grade >=3 TEAEs related to any study drug by SOC and PT, and by maximum grade;
- Summary of Treatment-emergent SAEs by SOC and PT;
- Summary of Treatment-emergent SAEs by SOC and PT, and by maximum CTCAE grade;
- Summary of Treatment-emergent SAEs related to study drug by SOC and PT, and by maximum CTCAE grade;

- Summary of Treatment-emergent SAEs related to any study drug by SOC and PT;
- Summary of TEAEs leading to dose reduction or interruption of any study drug by SOC and PT;
- Summary of treatment-related TEAEs leading to dose reduction or interruption of any study drug by SOC and PT;
- Summary of TEAEs leading to dose reduction of Rivoceranib / Sorafenib by SOC and PT;
- Summary of TEAEs leading to study drug interruption of any study drug by SOC and PT;
- Summary of TEAEs leading to interruption of SHR-1210 by SOC and PT;
- Summary of TEAEs leading to interruption of Rivoceranib/ Sorafenib by SOC and PT;
- Summary of TEAEs leading to dose reduction or interruption of Rivoceranib/Sorafenib by SOC and PT;
- Summary of TEAEs leading to discontinuation of any study drug by SOC and PT;
- Summary of treatment-related TEAEs leading to discontinuation of any study drug by SOC and PT;
- Summary of TEAEs leading to discontinuation of all study drugs by SOC and PT;
- Summary of treatment-related TEAEs leading to discontinuation of all study drugs by SOC and PT;
- Summary of TEAEs leading to discontinuation of SHR-1210 by SOC and PT;
- Summary of treatment-related TEAEs leading to discontinuation of SHR-1210 by SOC and PT;
- Summary of TEAEs leading to discontinuation of Rivoceranib/Sorafenib by SOC and PT;
- Summary of treatment-related TEAEs leading to discontinuation of Rivoceranib/Sorafenib by SOC and PT;
- Summary of TEAEs leading to death by SOC and PT;
- Summary of treatment-related TEAEs leading to death by SOC and PT;

For overall TEAE, TEAE with CTCAE Grade ≥3, TEAE excluding Hypertension, TEAE with CTCAE Grade ≥3 (excluding Hypertension), time-at-risk exposure-adjusted incidence rate will be summarized descriptively.

Time-at-risk (person-months at risk) is calculated as below:

- For subjects with events: (start date of the first event – first dose date +1) / 30.4375
- For subject without an event: (end date of on-treatment period– first dose date +1) / 30.4375

The time-at-risk exposure-adjusted incidence rate is calculated as below:

number of subjects with events / (total person-months at risk/ 100)

Listings of adverse events, serious adverse events, adverse events leading to study drug discontinuation, adverse events leading to dose reduction, adverse events leading to interruption, adverse events leading to death will be provided. The listings will include at least country/center ID, subject ID, age, sex, race, AE verbatim term, SOC and PT term, start/end date (study day), serious flag, CTC-AE grade, relationship to study drug, treatment for AEs, action taken, outcome. For SAE listing, reasons classified as serious will also be listed.

4.15.2.1.2. irAEs Assessed by Investigator

The number and percentage of subjects in each of the following categories for experimental group (SHR-1210 plus Rivoceranib) will be provided:

- Summary of irAE by SOC and PT, and by Maximum CTCAE grade;
- Summary of irAE leading to discontinuation of SHR-1210 by PT;
- Summary of irAE leading to discontinuation of Rivoceranib by PT;
- Summary of irAE leading to discontinuation of any study drug by PT;
- Summary of irAE leading to discontinuation of all study drugs by PT;
- Summary of serious irAE by PT.
- Summary of irAE leading to death by PT.

Treatment of irAEs Assessed by Investigator

- Number and percentage of subjects with at least one irAE treated with any concomitant medication for overall;
- Number and percentage of subjects with at least one irAE treated with any Systemically Administered Corticosteroids for overall;
- Number and percentage of subjects with at least one irAE treated with $\geqslant$ 40mg Systemically Administered Corticosteroids taken (total daily dose) for overall;

- Number and percentage of subjects with at least one irAE treated with any Immunosuppressive regimens for overall.

- Number and percentage of subjects with at least one irAE treated with any Hormone replacement therapy for overall.

Listing of irAE assessed by investigator will be provided. The listings will include at least country/center ID, subject ID, age, sex, race, AE verbatim term, SOC and PT term, date of first study drug administration, start/end date (study day), serious flag, CTC-AE grade, relationship to study drug, treatment for AE, Systemically Administered Corticosteroids Treatment, high dose systemic corticosteroids used, Immunosuppressive treatment, action taken, outcome.

4.15.2.1.3. Selected Adverse Events

The selected adverse events will be summarized by medical concept, PT and treatment group according to selected AEs for SHR-1210 and rivoceranib, respectively.

The number and percentage of subjects in each of the following categories by treatment group will be provided:

Selected Adverse Events for SHR-1210:

- Overview of selected AEs for SHR-1210 by treatment group;

- Summary of selected AEs for SHR-1210 by Medical concept and PT by treatment group;

- Summary of selected AEs for SHR-1210 related to any study drug by Medical concept and PT by treatment group;

- Summary of selected AEs for SHR-1210 by Medical concept and PT, and by maximum CTCAE grade by treatment group;

- Summary of selected AEs for SHR-1210 related to any study drug by Medical concept and PT, and by maximum CTCAE grade by treatment group;

- Summary of serious selected AEs for SHR-1210 by Medical concept and PT;

- Summary of serious selected AEs for SHR-1210 related to any study drug by Medical concept and PT;

- Summary of selected AEs for SHR-1210 leading to discontinuation of any study drug by Medical concept and PT;

- Summary of selected AEs for SHR-1210 leading to dose reduction or interruption of any study drug by Medical concept and PT;

- Summary of selected AEs for SHR-1210 leading to death by Medical concept and PT.

Selected Adverse Events for Rivoceranib:

- Overview of selected AEs for Rivoceranib by treatment group;
- Summary of selected AEs for Rivoceranib by Medical concept and PT;
- Summary of selected AEs for Rivoceranib related to any study drug by Medical concept and PT;
- Summary of selected AEs for Rivoceranib by Medical concept and PT, and by maximum CTCAE grade;
- Summary of selected AEs for Rivoceranib related to any study drug by Medical concept and PT, and by maximum CTCAE grade;
- Summary of serious selected AEs for Rivoceranib by Medical concept and PT;
- Summary of serious selected AEs for Rivoceranib related to any study by Medical concept and PT;
- Summary of selected AEs for Rivoceranib leading to discontinuation of any study drug by Medical concept and PT;
- Summary of selected AEs for Rivoceranib leading to dose reduction or interruption of any study drug by Medical concept and PT;
- Summary of selected AEs for Rivoceranib leading to death by Medical concept and PT.

Time to First Onset of Selected AEs for SHR-1210 / Rivoceranib

Time to first onset of selected AEs and grade 3 or higher selected AEs summarize by overall and Grade >=3 within medical concept category descriptively.

Treatment of Selected AEs for SHR-1210

- Number and percentage of subjects with at least one selected AEs for SHR-1210 treated with any concomitant medication by medical concept category;
- Summary of selected AEs for SHR-1210 requiring use of systemic Corticosteroids by medical concept and by maximum CTCAE grade;
- Summary of selected AEs for SHR-1210 requiring use of immunosuppressive treatment other than Corticosteroids by medical concept and by maximum CTCAE grade;
- Number and percentage of subjects with at least one selected AEs for SHR-1210 treated with $\geqslant$ 40mg systemically administered Corticosteroids taken (total daily dose) by medical concept category;

- Duration (total days) of systemically administered Corticosteroids by medical concept category for subjects treated with ⩾ 40mg/day systemically administered Corticosteroids.

- Duration (total days) of systemically administered Corticosteroids ⩾ 40mg/day by medical concept category.

- Number and percentage of subjects with at least one selected AEs Thyroid disorder and Type 1 Diabetes Mellitus for SHR-1210 treated with any Hormone replacement therapy by medical concept category and by maximum CTCAE grade;

Treatment of Selected AEs hepatotoxicity for Rivoceranib will also be analyzed with the same content as above:

- Number and percentage of subjects with at least one selected AEs hepatotoxicity for rivoceranib treated with any concomitant medication by medical concept category;
- Summary of selected AEs hepatotoxicity for rivoceranib requiring use of systemic Corticosteroids by medical concept and by maximum CTCAE grade;
- Summary of selected AEs hepatotoxicity for rivoceranib requiring use of immunosuppressive treatment other than Corticosteroids by medical concept and by maximum CTCAE grade;
- Number and percentage of subjects with at least one selected AEs hepatotoxicity for rivoceranib treated with ⩾ 40mg systemically administered Corticosteroids taken (total daily dose) by medical concept category;
- Duration (total days) of systemically administered Corticosteroids by medical concept category for subjects treated with ⩾ 40mg/day systemically administered Corticosteroids.

- Duration (total days) of systemically administered Corticosteroids ⩾ 40mg/day by medical concept category.

Resolution of Selected AEs for SHR-1210 / Rivoceranib

Time to resolution of selected AEs will be analyzed using K-M method when appropriate. The total number of selected AEs, number of selected AEs resolved, number of selected AEs ongoing will be provided. The minimum and maximum of time to resolution of selected AEs will be denoted by a "+".

For the subjects with selected AEs resolved, time to resolution of selected AEs will be summarized using mean, SD, median, minimum and maximum.

For the AE of reactive capillary endothelial proliferation (RCEP), all the summaries will include the following summary statistics:

- Location of RCEP

- Largest lesion numbers post baseline (<10, 10-19, 20-29, 30-39, 40-49, ≥50)
- Maximum diameter of lesions after baseline(mm)
- Complications
- Infection
- Ulcer
- Pain
- Sum of bleeding amount post baseline
- Any therapy (medications, laser, surgery, other)

A listing of selected adverse events for SHR-1210 or Rivoceranib, selected AEs for SHR-1210 requiring use of systemic Corticosteroids, selected AEs for SHR-1210 requiring use of immunosuppressive treatment other Than Corticosteroids will be provided with related to study drug, country/center ID, subject ID, age, sex, race, cluster, PT term, start/end date (study day), serious flag, CTC-AE grade, relationship to study drug, treatment for AE, action taken, outcome.

4.15.2.1.4. COVID-19 Related TEAEs

- Summary of COVID-19 related TEAEs by SOC and PT;
- Summary of COVID-19 related TEAEs leading to death by SOC and PT;
- Summary of COVID-19 related TEAEs leading to any study drug interruption or dose reduced by SOC and PT;
- Summary of COVID-19 related TEAEs leading to any study drug discontinuation by SOC and PT;

Listing of all COVID-19 related TEAEs will be provided. A complete list of COVID-19 related TEAEs is described in Appendix 2.

4.15.2.1.5. Deaths

Summaries of deaths will be provided with number and percentage of subjects by treatment group, categorized as following:

- All deaths (throughout the study)
- AE outcome of death only and onset date prior to initiation of new anti-cancer therapy.
- Reasons for death
  - Adverse Event

- Disease Progression
- Other

The date and reason of death will be provided in individual subjects listing together with study treatment received, date of first dose, date of last dose.

4.15.3. Laboratory Tests Results

Abnormal laboratory values will be flagged and identified as those above or below the normal reference range. The number and percentage of subjects experiencing clinically significant laboratory values at Baseline and post-baseline using the worst grade will be summarized by shift table. Similar analysis will be performed for fecal occult blood parameter, urinalysis, coagulation and HBV serology hepatitis B makers (e.g., HBsAg, HBeAg), HBV-DNA and HCV-RNA.

Laboratory parameters that are graded in NCI-CTCAE v4.03 will be summarized by NCI-CTCAE grade. The number and percentage of subjects experiencing toxicity at Baseline and post-baseline, using the worst grade, will be summarized by shift table. In the summary of laboratory parameters by NCI-CTCAE grade, parameters with NCI-CTCAE grading in both high and low directions will be summarized separately.

Liver function tests: Alanine aminotransferase (ALT), aspartate aminotransferase (AST), and total bilirubin (TBILI) are used to assess possible drug induced liver toxicity.

Summary of liver function tests will include the following categories. The number and percentage of subjects with each of the following categories after first dose recorded in eCRF will be summarized by group:

- ALT ⩾ 3×ULN, ALT ⩾ 5×ULN, ALT ⩾ 10×ULN, ALT ⩾ 20×ULN
- AST ⩾ 3×ULN, AST ⩾ 5×ULN, AST ⩾ 10×ULN, AST ⩾ 20×ULN
- (ALT or AST) ⩾ 3×ULN, (ALT or AST) ⩾ 5×ULN, (ALT or AST) ⩾ 10×ULN, (ALT or AST) ⩾ 20×ULN
- Total bilirubin (TBILI) ⩾ 2×ULN, ⩾ 5×ULN
- Potential Hy's law (refer to the definition listed in the section 3.3.2).

Categories will be cumulative, i.e., a subject with an elevation of AST ≥20×ULN will also appear in the categories ≥3×ULN, ≥5×ULN and ≥10×ULN.

AFP evaluated within 14 days of Screening period, Day 1 of each treatment cycle ⩾ 2, End-of-Treatment visit and SFU site visit will be listed.

Thyroid function, including thyroid stimulating hormone (TSH), FT3, FT4 will be listed.

A listing of all laboratory data will be provided. The listing includes abnormal laboratory values flagged and identified as those outside (above or below) the normal reference range, where appropriate.

A listing of subject's meeting laboratory criteria for potential Hy's law will be provided. The listing will include the visit, Meet potential Hy's Criteria, ALT, AST, TBIL, ALP.

4.15.4. ECOG-PS

Baseline and highest post-baseline ECOG-PS score (0-5) will be descriptively summarized using shift table by treatment group. A list of ECOG-PS assessments will be provided, which includes at least country/territorial ID, subject ID, date of assessment, study day and the scores.

4.15.5. Child-Pugh

Baseline and highest post-baseline Child-Pugh (Grade A, Grade B, Grade C) score will be descriptively summarized using shift table by treatment group. A list of Child-Pugh assessments will be provided, which includes at least country/territorial ID, subject ID, date of assessment, study day and the scores.

4.15.6. Vital Signs

The number and percentage of subjects with clinically significant abnormalities in vital signs at post-baseline visits will be provided for each parameter each criterion by treatment group. A listing of vital signs assessments will also be provided, which includes at least country/territorial, subject ID, date of assessment, study day, results and signs of clinically significant abnormalities. The detail of the criteria for clinically significant abnormalities is showed in Table 6.

4.15.7. Physical Examination

Physical examination collected throughout the study will be descriptively summarized by treatment group using a shift table and listed, including any clinical finding results. The shift table summarizes the changes in clinical assessment outcome as normal, abnormal without clinically significant, abnormal with clinically significant outcome between Baseline and worst post-baseline clinically abnormal result. If have OCT examinations, listing will be provided.

4.15.8. 12-Lead ECG

The overall clinical assessment outcomes are categorized as normal, abnormal without clinically significant, abnormal with clinically significant outcome. Baseline and worst post-baseline clinically abnormal result will be descriptively summarized using shift table.

Number and percentage of subjects with markedly abnormal value at post-Baseline will be provided for each treatment group for PR and QTcF. The detail of the criteria can be seen in Table 7.

4.15.9. Concentration-QTc Analysis for Rivoceranib

Concentration-QTc analysis will be performed based on the PKQTS. The post-dose rivoceranib plasma concentrations which are below quantifiable limit will be set to zero for the analysis.

A scatter plot of observed concentrations and QTcF changes from baseline (ΔQTcF) with a LOESS smooth line with 90% CI and a linear regression line will be provided for evaluation of the linearity of the concentration-QTc relationship.

4.15.10. Echocardiography

Echocardiography including left ventricular ejection fraction (LVEF) measured at investigator discretion during the study will be listed.

4.16. Interim Analyses

In this study, there is only one planned PFS analysis, which is planned to occur when ~332 PFS events are observed (expected at about 19 months). OS analyses will be conducted at the following time points: at final PFS analysis, at 251 (70%) deaths if there are less than 251 deaths at the final PFS analysis, and when ~359 OS events are observed (expected at about 36 months), respectively.

For the OS analysis, the O'Brien-Fleming alpha-spending function (OBF) will be used to adjust alpha except a nominal alpha of 0.001% will be spent at the time of the final PFS analysis if there are less than 251 deaths. Since the trial will not stop early if there are less than 251 deaths, the 0.001% is considered for administrative purposes to avoid trial stop early.

PFS

If SHR-1210 in combination with rivoceranib arm demonstrated superiority over sorafenib (i.e., one-sided p-value < 0.005), its corresponding alpha level of 0.005 one-sided will be re-allocated to test the OS hypothesis at 0.025 one-sided with updated efficacy boundaries per Lan-DeMets O'Brien-Fleming spending function (see Table 14).

OS

A formal interim analysis (IA) for OS will be conducted after 70% (~251) of the target OS events are observed at ~23 months. The OBF was used for calculating efficacy boundaries (original 0.020/updated 0.025 alpha level, one-sided). With ~251 OS events, the SHR-1210 in combination with rivoceranib mesylate will be considered superior than the sorafenib at the interim analysis, if the one-sided p-value for OS is less than 0.005/0.007, which corresponds to an observed HR of approximately 0.725/0.735. The trial will be considered positive at the final analysis, if the one-sided p-value for OS is less than 0.018/0.023, which corresponds to an observed HR of approximately 0.802/0.810. If the null hypothesis for OS is rejected at either the interim (70%) or the final, its corresponding alpha level of 0.020 one-sided will be re-allocated to PFS so that the hypothesis for PFS can be tested again at 0.025

one-sided. The boundary properties for the planned OS analyses based on the original and updated alpha levels can be found in Table 14.

However, in an anticipation that the number of OS events is < 251 OS events at the planned PFS analysis, a nominal alpha of 0.001% will be spent on testing OS at the time of the planned PFS analysis. Since the trial will not stop early if there are less than 251 deaths, the study will continue to the next planned OS interim analysis (70%) and could continue to final analysis after ~251 OS events and ~359 OS events are observed, respectively, according to the efficacy boundaries as described in Table 14.

4.16.1. Independent Data Monitoring Committee

The analysis of PFS and the formal interim analysis (IA) of OS will be conducted by a Data Monitoring Committee (DMC) that recommends early discontinuation of the study for unequivocal efficacy at the formal interim analysis (IA) of OS (~70% events), or continuation of the study for final analysis of OS. This committee includes independent members of oncologists, independent statistician and a chair person for each DMC meeting. The decision to adopt DMC recommendation lies with the Sponsor. Further detail regarding role/responsibility and conduct of the interim analyses by independent members can be found in the DMC charter. The exact analyses, including planned and supplementary analyses as requested by the independent members, are to be retained by the DMC members until study ends.

4.17. Multiplicity

The two primary endpoints for the study are OS and PFS (assessed by BIRC base on RECIST v1.1). There is only one planned PFS analysis, which is expected to occur when approximately 332 PFS events are observed at approximately 19 months. OS analyses will be conducted at the following time points: at final PFS analysis, at 251 (70%) deaths if there are less than 251 deaths at the final PFS analysis, and when ~359 OS events are observed at ~36 months, respectively.

For the OS analysis, the O'Brien-Fleming alpha-spending function (OBF) will be used to adjust alpha except a nominal alpha of 0.001% will be spent at the time of the final PFS analysis if there are less than 251 deaths. Since the trial will not stop early if there are less than 251 deaths, the 0.001% is considered for administrative purposes to avoid trial stop early.

The Type I error is strongly controlled using the Weighted-Holm procedure with α-reallocation approach between PFS and OS endpoints as described in Ye et al 2013 and O'Brien-Fleming alpha-spending function (OBF) within OS. The allocated α between two endpoints is α_1=0.005 (one-sided) for PFS and α_2=0.020 (one-sided) for OS. If PFS endpoint is significant at α_1=0.005, its α_1 level will be re-allocated to endpoint OS so that the hypothesis for OS can be tested again using its updated efficacy boundaries based on α level 0.025. Similarly, if endpoint OS is significant at α_2=0.020 according to the efficacy boundaries as determined by OBF, and PFS endpoint is not significant at α_1=0.005, its α_2 level will be re-allocated to PFS so that the hypothesis for PFS can be tested again at α level 0.025 (one-sided). In more

details, there are three possible scenarios of the testing results for PFS and OS as following:

1) PFS is significant

- If PFS is significant at α_1=0.005 and OS is significant at α level 0.005 (1-sided) at the time of interim analysis according to the efficacy boundaries as determined by the OBF alpha-spending function, alpha re-allocation will not be performed;

- If PFS is significant at α_1=0.005 and OS is not significant at α level 0.005 (1-sided) at the time of interim analysis according to the efficacy boundaries as determined by the OBF alpha-spending function, the α_1=0.005 will be re-allocated to endpoint OS, so that the hypothesis for OS can be tested using its updated efficacy boundary at α level 0.007 (1-sided), and the hypothesis for OS can be tested at α level 0.023 (1-sided) at the time of final analysis;

2) PFS is not significant and OS is significant

- If endpoint OS is significant at α level 0.005 (1-sided) at the time of formal interim analysis, and PFS endpoint is not significant at α_1=0.005, the total α_2=0.020 will be re-allocated to PFS so that the hypothesis for PFS can be tested again at α level 0.025 (one-sided).

- If endpoint OS is not significant at α level 0.005 (1-sided) at the time of formal interim analysis, but is significant at α level 0.018 (1-sided) at the time of final analysis, and PFS endpoint is not significant at α_1=0.005, the total α_2=0.020 will be re-allocated to PFS so that the hypothesis for PFS can be tested again at α level 0.025 (one-sided).

3) Both PFS and OS are not significant

- If PFS is not significant at α_1=0.005 and OS is neither significant at the time of interim analysis nor at final analysis at 0.020 according to the efficacy boundaries as determined by the OBF alpha-spending function, alpha re-allocation will not be performed;

The boundary properties for the planned OS analyses based on the original and updated alpha levels are summarized in Table 14, below:

Table 14 Boundary Properties for Planned Analyses of OS

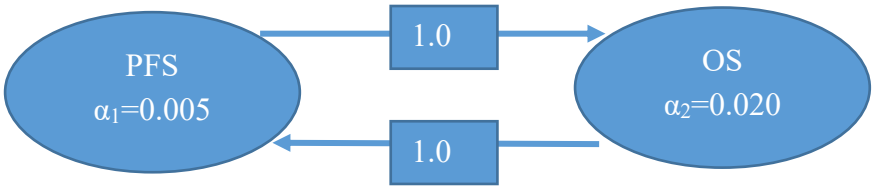
Analysis	Value	α=0.020	α=0.025
Interim analysis：~70%	Z	-2.549	-2.440
Events：251	p (one-sided)	0.005	0.007

Months：~23	HR	0.725	0.735
Final analysis	Z	-2.089	-2.000
Events：359	p (one-sided)	0.018	0.023
Months：~36	HR	0.802	0.810

Note: Results were generated using the EAST® 6.4.1. software; HR=Hazard ratio (SHR-1210 + Rivoceranib/Sorafenib) less than 1 indicates SHR-1210+rivoceranib group is superior than Sorafenib arm. The one-sided nominal significance level and stopping boundary will be determined by the actual percentage of events when the interim analysis is conducted.

The overall type I error will be controlled under 0.025 one sided. If both PFS and OS endpoints are not significant, then α-reallocation will not perform. If both the PFS and OS endpoints are significant, ORR assessed by BIRC according to RECIST v1.1 at the time of 70% OS events observed will be tested at α level of 0.025. The graphical approach in Figure 1 shows the α-reallocation strategy for endpoints PFS and OS:

Figure 1: Type I error reallocation strategy for endpoints PFS and OS.



Above shows the initial one-sided α-allocation for each hypothesis in the ellipse representing the hypothesis. The weights for re-allocation from each hypothesis to another are represented in the rectangular boxes on the lines connecting hypotheses.

However, if there are less than 251 OS events at the planned PFS analysis at ~19 months, a nominal α (0.001%, independent of OBF approach to allocate alpha) will be used to test OS, but the trial will not stop based on a positive OS result at the time of the PFS analysis. The study will continue to the next planned OS interim analysis (70%) and could continue to the final analysis after ~251 OS events and ~359 OS events at ~23 and ~36 months, respectively, at the alpha level as determined by the OBF boundaries and the actual number of OS events observed. The nominal alpha (0.001%) is considered, as this has negligible impact on overall Type I error rate and the study will not stop early for a positive OS result at the time of the PFS analysis.

As cut-off date of 2022-02-08 for the formal OS interim analysis, 262(73%) OS events are expected to be observed. A nominal α of 0.0087 using the Lan-DeMets procedure with an O'Brien-Fleming boundary is planned to be used to test OS. If one-sided p-value obtained from the stratified log-rank test is < 0.0087, superiority of SHR-1210 plus rivoceranib over sorafenib in OS will be demonstrated.

4.18. Subgroup Analyses

To determine whether the treatment effect is consistent across various subgroups, HR and 95% CI for the primary endpoints OSand PFS (assessed by BIRC base on RECIST v1.1) will be estimated using unstratified Cox proportional hazard model, key secondary endpoint ORR (assessed by BIRC base on RECIST v1.1) and 95%CI will be calculated, and plotted within each category of the following classification variables according to eCRF data:

- Age (<65 years vs. ≥65 years);
- Gender (male vs. female);
- Geographical region (Asia vs. non-Asia countries vs Mainland China/HongKong/Taiwan);
- Race (Asian vs. White vs. African American vs. others);
- BCLC stage (Stage B vs. Stage C);
- Macrovascular invasion and/or extrahepatic metastasis (Presence vs. Absence);
- Macrovascular Vascular invasion (Yes vs. No);
- Extrahepatic metastasis (Yes vs. No);
- Root cause of HCC (HBV vs. HCV vs. Non-viral);
- AFP elevation at the baseline (<400ng/mL vs ≥400ng/mL);
- Involved disease organs assessed by BIRC (≤1 vs. 2 vs. ≥3);
- Previous local therapy (yes vs. no);
- Weight (＜60 kg vs. ≥60 kg);
- ECOG-PS Score (0 vs. 1);
- PD-L1 category 1 (TPS <1%, TPS ≥1%, unknown)
- PD-L1 category 2 (TPS <5%, TPS ≥5%, unknown)
- PD-L1 category 3 (CPS <1, CPS ≥1, unknown)
- PD-L1 category 4 (CPS <5, CPS≥5, unknown)

Forest plots of HRs and 95% CIs will be provided for each of the primary endpoints. Subgroups with small sample size may reflect in wide confidence intervals, and point estimates that may not be estimable will be presented as NE=Not Estimable, or indicated otherwise. If any of the analyses was not carried out because the size of any category was too small, it will be noted in the relevant section(s) of the CSR.

For Geographical region (Asia vs. non-Asia countries vs Mainland China/HongKong/Taiwan), demographic and other baseline characteristics will also be summarized. For Geographical region of Mainland China/HongKong/Taiwan, the HR and 95% CI for the primary endpoint OS and PFS (assessed by BIRC base on RECIST v1.1) will also be estimated using stratified Cox proportional hazard model, the treatment comparison will be based on a stratified Log-rank test.

5. REFERENCES

[1] Amit O, et al. Blinded independent central review of progression in cancer clinical trials: Results from a meta-analysis and recommendation from a PhRMA working group. European Journal of Cancer 47:1772-1778, 2011.

[2] Fayers P, Aaronson NK, Bjordal K, Curran D, Groenvold M on behalf of the EORTC Quality of Life Study Group. EORTC QLQ-C30 Scoring Manual: 2nd Edition 1999. Available on request.

[3] Osoba_D, Rodrigues G, Myles J ZEE B, PATER J, Interpreting the significance of changes in health-related quality-of-life scores. 1998; J Clin Oncol 16 139-144.

[4] Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: An EQ-5D-5L value set for England, Health Economics. 2018; 27:7–22.

[5] Robins, J. and Finkelstein, D. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with Inverse Probability of Censoring Weighted (IPCW) Log-Rank Tests. Biometrics 2000; 56: 779-788.

[6] Latimer, N., Abrams, K., Lambert, P., Crowther, M., et al. Adjusting for treatment switching in randomised controlled trials – A simulation study and a simplified two-stage method. Statistical Methods in Medical Research 2017; 26(2): 724-751.

[7] Ye, Yining, et al. "A group sequential Holm procedure with multiple primary endpoints." Statistics in medicine 32.7 (2013): 1112-1124.

6. APPENDIX

Appendix 1. Selected Adverse Events defined by Sponsor

Selected Adverse Events for SHR-1210:

Category	SMQs/ PTs
Pneumonitis (23 PTs)	PTs: Acute interstitial pneumonitis, Alveolar lung disease, Alveolitis, Autoimmune lung disease, Bronchiolitis, Diffuse alveolar damage, Eosinophilic pneumonia, Eosinophilic pneumonia acute, Eosinophilic pneumonia chronic, Hypersensitivity pneumonitis, Immune-mediated lung disease, Interstitial lung disease, Lung opacity, Obliterative bronchiolitis, Organising pneumonia, Pneumonitis, Pulmonary fibrosis, Pulmonary toxicity, Pulmonary vasculitis, Idiopathic pneumonia syndrome, Idiopathic interstitial pneumonia, Lung infiltration, Pulmonary sarcoidosis
Colitis （22 PTs）	HLT Colitis (excl infective): Acute haemorrhagic ulcerative colitis, Allergic colitis, Autoimmune colitis, Colitis, Colitis erosive, Colitis ischaemic, Colitis microscopic, Colitis ulcerative, Crohn's disease, Enterocolitis haemorrhagic, Immune-mediated enterocolitis, Inflammatory bowel disease, Necrotising colitis, Terminal ileitis PT: Enterocolitis, Proctitis haemorrhagic, Proctitis, Diarrhoea haemorrhagic, Frequent bowel movements, Enteritis, Vasculitis gastrointestinal, Diarrhoea
Thyroid disorder：	
Hypothyroidism (10 PTs)	HLT Thyroid hypofunction disorders: Autoimmune hypothyroidism, Hypothyroidic goitre, Hypothyroidism, Immune-mediated hypothyroidism, Myxoedema, Myxoedema coma, Primary hypothyroidism, Secondary hypothyroidism, Tertiary hypothyroidism, Thyroid atrophy
Hyperthyroidism (14 PTs)	HLT Thyroid hyperfunction disorders: Basedow's disease, Endocrine ophthalmopathy, Exophthalmos, Hashitoxicosis, Hyperthyroidism, Immune-mediated hyperthyroidism, Marine Lenhart syndrome, Primary hyperthyroidism, Secondary hyperthyroidism, Thyroid dermatopathy, Thyrotoxic crisis, Thyrotoxic periodic paralysis, Toxic goitre, Toxic nodular goitre
Thyroiditis (10 PTs)	HLT Acute and chronic thyroiditis: Atrophic thyroiditis, Autoimmune thyroiditis, Immune-mediated thyroiditis, Silent thyroiditis, Thyroiditis, Thyroiditis acute, Thyroiditis chronic, Thyroiditis fibrous chronic, Thyroiditis subacute PTs: Thyroid disorder
Adrenal Insufficiency (14 PTs)	HLT Adrenal cortical hypofunctions: Addison's disease, Adrenal androgen deficiency, Adrenal atrophy, Adrenal insufficiency, Adrenal suppression, Adrenocortical insufficiency acute, Glucocorticoid deficiency, Hypoaldosteronism, Mineralocorticoid deficiency, Primary adrenal insufficiency, Secondary adrenocortical insufficiency, Steroid withdrawal syndrome, Immune-mediated adrenal insufficiency, Blood corticotrophin increased
Type 1 Diabetes Mellitus (11 PTs)	PTs: Diabetes mellitus, Diabetic ketoacidosis, Fulminant type 1 diabetes mellitus, Hyperglycaemia, Ketoacidosis, Latent autoimmune diabetes in adults, Type 1 diabetes mellitus, Euglycaemic diabetic ketoacidosis, Diabetic ketoacidotic hyperglycaemic coma, Blood glucose increased, Ketosis-prone diabetes mellitus
Hypophysitis (8 PTs)	PTs: Hypophysitis, Hypopituitarism, Lymphocytic hypophysitis, Immune-mediated hypophysitis, Hypothalamic pituitary adrenal axis

	suppression, Hypothalamo-pituitary disorder, Adrenocorticotropic hormone deficiency, Blood corticotrophin decreased
Nephritis and renal dysfunction (39 PTs)	HLT Nephritis NEC: Autoimmune nephritis, Immune-mediated nephritis, Lupus nephritis, Nephritis, Tubulointerstitial nephritis, Nephritis haemorrhagic HLT Glomerulonephritis and nephrotic syndrome: Glomerulonephritis, Glomerulonephritis acute, Glomerulonephritis membranous, IgA nephropathy, Nephritis allergic, Nephrotic syndrome, Glomerulonephritis membranoproliferative, Glomerulonephritis minimal lesion, Glomerulonephritis proliferative, Glomerulonephritis rapidly progressive, Chronic autoimmune glomerulonephritis, Mesangioproliferative glomerulonephritis, Fibrillary glomerulonephritis, Focal segmental glomerulosclerosis HLT Renal failure and impairment: Acute kidney injury, Renal failure, Renal impairment, Renal injury, Oliguria, Anuria PTs: Renal disorder, Renal tubular disorder, Renal tubular acidosis, Renal tubular dysfunction, Renal tubular atrophy, Renal tubular injury, Renal tubular necrosis, Nephropathy toxic, Azotaemia, Creatinine renal clearance decreased, Glomerular filtration rate decreased, Blood creatinine increased, Hypercreatininaemia
Skin reactions (99 PTs)	SMQ Severe cutaneous adverse reactions (Narrow search) HLT Rashes, eruptions and exanthems NEC: Genital rash, Nodular rash, Rash, Rash erythematous, Rash macular, Rash maculo-papular, Rash papular, Rash pruritic, Rash vesicular, Administration site rash, Anal rash, Application site rash, Butterfly rash, Catheter site rash, Eyelid rash, Heliotrope rash, Implant site rash, Incision site rash, Infusion site rash, Injection site rash, Instillation site rash, Medical device site rash, Mucocutaneous rash, Paraneoplastic rash, Penile rash, Perineal rash, Rash maculovesicular, Rash morbilliform, Rash neonatal, Rash rubelliform, Rash scarlatiniform, Stoma site rash, Systemic lupus erythematosus rash, Vaccination site rash, Vascular access site rash, Vessel puncture site rash, Viral rash, Vulvovaginal rash HLT Bullous conditions: Autoimmune blistering disease, Blister, Bullous impetigo, Dermatitis herpetiformis, Epidermolysis bullosa, Pemphigoid, Pemphigus HLT Dermatitis and eczema: Autoimmune dermatitis, Dermatitis, Dermatitis allergic, Dermatitis atopic, Dermatitis contact, Eczema, Eczema infected, Eczema vesicular, Hand dermatitis, Immune-mediated dermatitis, Seborrhoeic dermatitis, Urticarial dermatitis PTs: Dermatitis acneiform, Drug eruption, Erythema, Keratolysis exfoliativa acquired, Macule, Papule, Pruritus, Pruritus genital, Skin disorder, Skin exfoliation, Skin lesion, Skin reaction, Skin toxicity, Solar dermatitis, Urticaria, Rash papulosquamous, Rash follicular, Rash pustular, Southern tick-associated rash illness, Candida nappy rash, Septic rash, Vasculitic rash, Lichenoid keratosis
Pancreatitis (12 PTs)	PTs: Immune-mediated pancreatitis, Pancreatitis, Pancreatitis acute, Haemorrhagic necrotic pancreatitis, Pancreatitis necrotising, Amylase increased, Pancreatitis relapsing, Lipase increased, Autoimmune pancreatitis, Hyperamylasaemia, Hyperlipasaemia, Pancreatitis haemorrhagic
Nervous system toxicity (60 PTs)	PTs: Autoimmune demyelinating disease, Chronic inflammatory demyelinating polyradiculoneuropathy, Clinically isolated syndrome, Concentric sclerosis, Demyelinating polyneuropathy, Leukoencephalomyelitis, Myelitis transverse, Neuromyelitis optica pseudo relapse, Neuromyelitis optica spectrum disorder, Noninfectious myelitis, Noninfective encephalomyelitis, Subacute inflammatory demyelinating polyneuropathy, Acute flaccid myelitis, Axonal and

	demyelinating polyneuropathy, Demyelination, Encephalomyelitis, Myelitis, Hypoglossal nerve paresis, IIIrd nerve paresis, IVth nerve paresis, Paresis cranial nerve, Trigeminal nerve paresis, VIth nerve paresis, Optic neuritis, Encephalitis toxic, Subacute sclerosing panencephalitis, Panencephalitis, Encephalitis haemorrhagic, Mononeuritis, Neuropathy peripheral, Immune-mediated neurological disorder, Multifocal motor neuropathy, Acute polyneuropathy, Rasmussen encephalitis, Meningitis aseptic, Encephalitis lethargica, Guillain-Barre syndrome, Polyneuropathy idiopathic progressive, Lupus encephalitis, Neurotoxicity, Neuritis, Neuromyopathy, Neuromuscular toxicity, Miller fisher syndrome, Myasthenic syndrome, Encephalitis brain stem, Encephalitis, Meningitis, Autonomic neuropathy, Autoimmune neuropathy, Encephalitis autoimmune, Axonal neuropathy, Limbic encephalitis, Encephalitis allergic, Myasthenia gravis, Myasthenia gravis crisis, Noninfective encephalitis, Facial nerve disorder, Immune-mediated encephalitis, Nerve injury
Ocular toxicity (19 PTs)	PTs: Visual impairment, Blindness, Episcleritis, Scleritis, Vogt-Koyanagi-Harada disease, Ocular myasthenia, Ocular sarcoidosis, Blepharitis, Cyclitis, Conjunctivitis, Autoimmune uveitis, Uveitis, Iritis, Iridocyclitis, Papilloedema, Retinal detachment, Chorioretinitis, Ulcerative keratitis, Keratitis
Rheumatoid/ skeletal muscle toxicity (28 PTs)	PTs: Polymyalgia rheumatica, Immune-mediated myositis, Arthritis, Arthralgia, Joint swelling, Paraneoplastic dermatomyositis, Inclusion body myositis, Eosinophilia myalgia syndrome, Necrotising myositis, Polymyositis, Myositis ossificans circumscripta, Infective myositis, Rhabdomyolysis, Synovitis, Viral myositis, Dermatomyositis, Myositis-like syndrome, Rheumatoid arthritis, Sarcoidosis, Myositis, Myopathy, Myalgia, Pain in extremity, Pyomyositis, Myositis ossificans, Arthropathy, Immune-mediated arthritis, Sjogren's syndrome
Myocarditis (11 PTs)	PTs: Myocarditis, Myocarditis infectious, Radiation myocarditis, Lupus myocarditis, Viral myocarditis, Myocarditis mycotic, Myocarditis bacterial, Myocarditis septic, Autoimmune myocarditis, Hypersensitivity myocarditis, Immune-mediated myocarditis
Reactive Capillary Endothelial Proliferation	PT: Reactive capillary endothelial proliferation
Infusion-related reaction	PTs: Infusion related hypersensitivity reaction, Infusion related reaction, Hypersensitivity, Drug hypersensitivity, Type I hypersensitivity, Anaphylactoid reaction, Anaphylactoid shock, Anaphylactic reaction, Anaphylactic shock, Cytokine release syndrome （only include the select events with an onset date within one day of camrelizumab infusion: start date of AE – injection date of SHR1210<=1）

Selected Adverse Events for Rivoceranib:

Category	SMQs/ PTs
Hypertension	SMQ: Hypertension (Narrow search)
Proteinuria	SMQ: Proteinuria (Narrow search)
Palmar-plantar erythrodysaesthesia syndrome	PT: Palmar-plantar erythrodysaesthesia syndrome
Hemorrhage	SMQ: Haemorrhage terms (excl laboratory terms)
Hepatotoxicity	SMQ: Hepatitis, non-infectious (Narrow search)

	SMQ: Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (Narrow search) SMQ: Cholestasis and jaundice of hepatic origin (Narrow search) SMQ: Liver related investigations, signs and symptoms (Narrow search)
QT prolongation	SMQ: Torsade de pointes/QT prolongation (Narrow search)
Gastrointestinal perforation	SMQ: Gastrointestinal perforation
Thromboembolic events	SMQ: Embolic and thrombotic events
Arterial thromboembolic events	SMQ: Embolic and thrombotic events, arteria
Venous thromboembolic events	SMQ: Embolic and thrombotic events, venous
Posterior reversible encephalopathy syndrome	PTs: Posterior reversible encephalopathy syndrome, Leukoencephalopathy
Wound healing impaired	PTs: Impaired healing, Incisional hernia, Incision site impaired healing, Postoperative wound complication

Appendix 2. COVID-19 related TEAE

SMQ	PTs
COVID-19	PTs: Asymptomatic COVID-19, Congenital COVID-19，Coronavirus infection，Coronavirus test positive，COVID-19，COVID-19 immunisation，COVID-19 pneumonia，COVID-19 prophylaxis，COVID-19 treatment，Exposure to SARS-CoV-2，Multisystem inflammatory syndrome in children，Occupational exposure to SARS-CoV-2，Post-acute COVID-19 syndrome，SARS-CoV-2 antibody test positive，SARS-CoV-2 carrier，SARS-CoV-2 RNA decreased，SARS-CoV-2 RNA fluctuation，SARS-CoV-2 RNA increased，SARS-CoV-2 sepsis，SARS-CoV-2 test false negative，SARS-CoV-2 test positive，SARS-CoV-2 viraemia，Suspected COVID-19，Vaccine derived SARS-CoV-2 infection

Appendix 3. SAS code for primary analysis of PFS and OS

- Kaplan-Meier Method and stratified Log-Rank test：

```
Proc lifetest data=dataset_name;
Time PFSTIME*censor (1);   /***1 means censored***/
Strata strata1 strata2 strata3/Group=TREATMENT;
Run;
```

- Stratified Cox Proportional Hazards Model：

```
Proc phreg data=dataset_name (where= (TREATMENT in (1 2)));
Class TREATMENT (ref='2');   /***1= Experimental group，2=Control group***/
model PFSTIME*censor(1)= TREATMENT/ties=discrete; /***1 means censored ***/
Strata strata1 strata2 strata3;
Hazardratio TREATMENT/diff=ref;
Run;
```

Strata1, strata2 and strata3 means stratification factors, which are macrovascular invasion and/or extrahepatic metastasis (presence vs. absence), geographical region (Asia vs. countries outside of Asia) and Baseline AFP (AFP < 400 ng/mL vs. AFP ⩾ 400 ng/mL).