Title: A Randomized, Double-Blind, Placebo-Controlled, Multicenter Phase III Clinical Trial of Apatinib Mesylate Tablets as Second-Line Treatment in Patients with Advanced Hepatocellular Carcinoma
Clinical trial registration number: NCT02329860
Version Date: 18 Jul. 2019

A Randomized, Double-Blind, Placebo-Controlled, Multicenter Phase III Clinical Trial of Apatinib Mesylate Tablets as Second-Line Treatment in Patients with Advanced Hepatocellular Carcinoma

Protocol No.: APTN-III-HCC

Statistical Analysis Plan

(V2.0)

Sponsors: Jiangsu Hengrui Pharmaceuticals Co., Ltd.
Shanghai Hengrui Pharmaceutical Co., Ltd.

Statistical Institution: Department of Biostatistics, School of Public Health, Nanjing Medical University

18 Jul. 2019

Statistical Analysis Plan
(V2.0)

Investigational Drug:	Apatinib mesylate tablets
Trial Title:	A Phase III Clinical Trial of Apatinib Mesylate Tablets as Second-Line Treatment in Patients with Advanced Hepatocellular Carcinoma
Clinical Phase:	Phase III
Sponsor:	Jiangsu Hengrui Pharmaceuticals Co., Ltd., Shanghai Hengrui Pharmaceutical Co., Ltd.
Principal Investigator:	[redacted]
[redacted]	[redacted]
Written by:	[redacted]

Department of Biostatistics, School of Public Health, Nanjing Medical University

Signature Page of Statistical Analysis Plan

Trial Title: A Phase III Clinical Trial of Apatinib Mesylate Tablets as Second-Line Treatment in Patients with Advanced Hepatocellular Carcinoma

Principal Investigator:

Signature_____ Date _____

Sponsors: Jiangsu Hengrui Pharmaceuticals Co., Ltd.
Shanghai Hengrui Pharmaceutical Co., Ltd.

Signature_____ Date _____

Statistical Institution: Department of Biostatistics, School of Public Health, Nanjing Medical University

Written by_____ Date _____

Reviewed by _____ Date _____

TABLE OF CONTENTS

1 ABBREVIATIONS 1
2 PROTOCOL SYNOPSIS 2
2.1 Study Objectives 2
2.2 Study Population 2
2.3 Study Design 2
2.4 Sample Size 2
2.5 Randomization Method 3
2.6 Interim Analysis 3
3 EFFICACY EVALUATION ENDPOINTS 3
3.1 Primary Efficacy Endpoint 3
3.2 Secondary Efficacy Endpoints 3
4 SAFETY ENDPOINTS 4
4.1 Adverse Events 4
4.2 Vital Signs and Physical Examination 5
4.3 Laboratory Tests 5
5 OTHER EVALUATION PARAMETERS 6
6 ANALYSIS SETS 6
6.1 Full Analysis Set (FAS) 6
6.2 Per-Protocol Set (PPS) 6
6.3 Safe Set (SS) 6
7 STATISTICAL ANALYSIS 7
7.1 General Principles 7
7.1.1 Significance level 7
7.1.2 Hypothesis test 7
7.1.3 Data principles 7
7.2 Study Population 9
7.2.1 Case Distribution 9
7.2.2 Protocol violations and deviations 9
7.2.3 Analysis of subject characteristics of two groups 10

7.3 Efficacy Evaluation 10
7.3.1 Analysis of primary efficacy endpoint 10
7.3.2 Analysis of secondary efficacy endpoints 11
7.4 Safety Evaluation 12
7.4.1 Adverse events 12
7.4.2 Drug exposure and dose modification 13
7.4.3 Vital signs 13
7.4.4 Laboratory test parameters 13
7.4.5 ECG 13
7.4.6 Physical examination 13
7.4.7 Concomitant medications 13
7.5 Interim Analysis 14
7.6 Other Analyses 14
7.6.1 Subgroup analysis 14
7.6.2 Other Analyses 14
8 STATISTICAL ANALYSIS SOFTWARE 15
9 REFERENCES 16
10 ATTACHED TABLES 17
Appendix 1. Classification table of protocol deviations 17
Appendix 2. Adverse events of special interest 19
Appendix 3. Derivative variables for apatinib/placebo 20

1 ABBREVIATIONS

AE	Adverse Event
CI	Confidence Interval
CMH	Cochran–Mantel–Haenszel
NCI-CTCAE 4.0	The National Cancer Institute-Common Terminology Criteria for Adverse Events V4.0
DCR	Disease Control Rate
DOR	Duration of Response
EORTC QLQ-C30	Quality of Life Questionnaire
FAS	Full Analysis Set
HCC	Hepatocellular Carcinoma
HR	Hazard Ratio
ITT	Intention to Treat
M	Median
Max	Maximum
Mean	Mean
MedDRA	Medical Dictionary for Regulatory Activities
Min	Minimum
ORR	Objective Response Rate
OS	Overall Survival
PFS	Progression-Free Survival
PPS	Per-Protocol Set
Q_1	25^{th} Percentile
Q_3	75^{th} Percentile
SAE	Serious Adverse Event
Std	Standard Deviation
SS	Safety Set
TTP	Time to Progression
ULN	Upper Limit of Normal

2 PROTOCOL SYNOPSIS

2.1 Study Objectives

To observe and evaluate the efficacy and safety of apatinib mesylate tablets (hereinafter referred to as apatinib) as second-line treatment in patients with advanced hepatocellular carcinoma (HCC).

2.2 Study Population

Patients with relapsed or metastatic HCC refractory or intolerant to systemic chemotherapy (oxaliplatin alone or in a therapeutic combination) and/or molecular targeted therapy with sorafenib.

2.3 Study Design

A randomized, double-blind, placebo-controlled, and multicenter trial design.

2.4 Sample Size

In this trial, placebo is used as a control and a superiority test is performed between the two groups.

Assuming the median OS for the placebo group and the apatinib group is 6 months and 8.5 months, respectively, according to FAS analysis results for a phase II clinical trial and relevant literature reports, assuming the two-sided $\alpha = 0.05$ and 80% power, with 2:1 randomization (treatment vs placebo), the lost to follow-up rate for each group is 15%, and assuming the accrual period is 12 months, and the duration of study period is 36 months, 331 subjects is required to be enrolled (222 subjects in the treatment group, 109 subjects in the placebo group) according to the calculation formula of the Log-rank test for the OS comparison between two groups (calculated by the PASS software), see Table 1 below. Considering dropout and other factors, this trial plans to enroll 360 patients, including 240 for the treatment group and 120 for the placebo group. According to the resolution of scientific committee on 31 Oct. 2016, 390 subjects are planned to be enrolled (see the Resolution of Scientific Committee on 31 Oct., 2016).

Table 1. Results of sample size estimation

Parameter	Type I Error	Median Survival		Sample Size Ratio Treatment:Control	Results		
		Treatment Group	Placebo Group		Treatment Group	Placebo Group	Total Number of Subjects
OS	0.05	8.5	6	2:1	222	109	331

2.5 Randomization Method

The randomization method used in this trial is as follows: Subjects are randomized to two groups in a 2:1 ratio to receive apatinib or placebo treatment.

The subjects are allocated by central randomization and enrolled competitively at all sites. The central randomization system provided by the Department of Biostatistics of Nanjing Medical University will be used for central randomization. After all eligible subjects are screened by study personnel at each study site participating in this trial and confirmed by the investigator of each site, their screening information will be entered into the randomization system to generate the randomization numbers and drug numbers. Corresponding study drugs will be dispensed based on the drug numbers. Subjects must begin the assigned study treatment within 48 h after randomization.

Stratification factors for randomization include:

- ECOG PS: 0 or 1
- Treatment with sorafenib: yes or no
- Presence of vascular invasion and/or extrahepatic metastasis: yes or no

2.6 Interim Analysis

None.

3 EFFICACY EVALUATION ENDPOINTS

3.1 Primary Efficacy Endpoint

Overall survival (OS): the time from the date of randomization to death of any cause.

3.2 Secondary Efficacy Endpoints

- Progression free survival (PFS): the time from the date of randomization to onset of progressive disease (PD) (as per RECIST 1.1) or death, whichever occurs first;
- 3-, 6-, and 12-month PFS rates;
- Time to progression (TTP): the time from the date of randomization to any documentation of radiologic progression (as per RECIST 1.1);
- Duration of response (DOR): the time from the first recording of tumor response to the first recording of objective tumor progression or death of any cause;
- Objective response rate (ORR): the proportion of subjects with confirmed complete response (CR) and partial response (PR) when the objective response is assessed as per RECIST 1.1;

- The proportion of subjects with stable disease (SD) for ≥ 6 weeks;
- Disease control rate (DCR): the proportion of subjects with confirmed CR, PR, and SD (for ≥ 6 weeks);
- 6- and 12-month survival rates;
- Quality of life score: EORTC QLQ-C30 and HCC18.

4 SAFETY ENDPOINTS

Safety evaluation parameters include vital signs, laboratory measurements, adverse events (AEs), serious adverse events (SAEs), treatment-related AEs and SAEs, and AEs specific to this type of drugs (e.g., hypertension, proteinuria, and hand-and-foot syndrome); judged per NCI-CTCAE V4.0.

4.1 Adverse Events

AEs, treatment-related adverse events (TRAEs), and SAEs that have occurred during the study period.

An AE refers to any untoward medical occurrence in a clinical investigation subject and which does not necessarily have a causal relationship with the treatment. AEs include but are not limited to: laboratory abnormalities; clinically significant symptoms and signs; allergies;

According to the needs of management, safety monitoring (reporting of AEs or SAEs) should be performed from the signing of the informed consent form by the subject to 28 days after the last dose.

Treatment-related AEs (TRAEs): All toxicities and unintentional reactions to a drug associated with any dose should be considered as Treatment-related AEs (TRAEs). Treatment-related AE means that a potential causal relationship between the drug and AE is at least reasonable, which means that the relationship cannot be excluded. The potential relationship between an AE and the investigational drug is assessed with 5 grades of causality: definitely related, probably related, possibly related, unlikely related, and not related. Events that are assessed to be "definitely related", "probably related", and "possibly related" are considered related to the investigational drug. When calculating the incidence of Treatment-related AEs, the total number of these three categories is used as the numerator and the total number of subjects in safety set is used as the denominator.

A SAE refers to a medical occurrence during the clinical trial that results in hospitalization, prolonged hospitalization, disability, incapacity, life threatening events or death, or congenital malformation. The following unexpected medical events are included:

- Events resulting in death;

- Life-threatening events (defined as events where the subject is at immediate risk of death at the time of the onset);
- Events resulting in hospitalization or prolonged hospitalization;
- Events resulting in permanent or serious disability/incapacity;
- Congenital anomalies or birth defects;
- Overdose;
- Pregnancy during the clinical trial.

Criteria for other SAEs specified in the protocol:

- Cardiac insufficiency — defined as any symptom or sign of reduced LVEF (NCI CTCAE Grade 3), or LVEF reduced by ≥ 20% from baseline and below the lower limit of normal (or reduced to < 50%).
- ALT > 5 × ULN and total bilirubin > 2 × ULN (such as direct bilirubin > 35%; the bilirubin ratio should be measured).

4.2 Vital Signs and Physical Examination

- Vital signs: body temperature, heart rate, respiratory rate, and blood pressure;
- Physical examination: including general conditions, head and face, skin, lymphatic system, eyes (sclera, pupil), respiratory system, cardiovascular system, abdomen (including liver and spleen), reproductive-urinary system, neuromuscular system, nervous system, mental status, and other conditions.

4.3 Laboratory Tests

- Hematology: including hemoglobin (HB), red blood cell count (RBC), white blood cell count (WBC), absolute neutrophil count (ANC), neutrophil ratio (N), lymphocyte ratio (Lym), monocyte ratio (Mon), and platelet count (PLT);
- Urinalysis: including urine protein (PRO), urine red blood cells (RBC), and urine white blood cells (WBC);
- Stool routine: occult blood;
- Liver and kidney functions: including total bilirubin (Tbil), direct bilirubin (dBiL), indirect bilirubin (iBiL), alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALP), Gamma-glutamyltransferase (GGT), lactate dehydrogenase (LDH), total protein (TP), albumin (ALB), creatinine (Cr), urea nitrogen (BUN), and uric acid (UA);

- Electrolytes: potassium ion (K^+), sodium ion (Na^+), chloride ion (Cl^-), calcium ion (Ga^{2+}), and blood phosphorus (P);Coagulation function: prothrombin time (PT), activated partial thromboplastin time (APTT), thrombin time (TT), plasma fibrinogen (Fbg), and international normalized ratio (INR);
- 12-Lead ECG: QTc value, heart rate, QTcF, and ECG description.

5 OTHER EVALUATION PARAMETERS

- Laboratory parameters of tumor markers: alpha-fetoprotein (AFP).

6 ANALYSIS SETS

There are three analysis sets in this study: full analysis set (FAS), per-protocol set (PPS), and safety set (SS).

6.1 Full Analysis Set (FAS)

According to the intention-to-treat (ITT) principle, the FAS includes all the randomized subjects who have received at least one dose of the study drug.

According to the ITT principle, all the subjects will be included in the analysis according to their randomized groups. The FAS is the primary analysis set for the efficacy analysis of this study.

6.2 Per-Protocol Set (PPS)

The PPS is a subset of the FAS and is defined as all subjects who are in compliance with the protocol, have good compliance, have received at least 1 cycle of the study drug (except for those with clear medical evidence of PD), and have completed the eCRF.

The statistical analysis will be performed for the drug efficacy based on the FAS and PPS. Before database locking, the principal investigator, statistician, and sponsor determine the final PPS during the data review meeting.

6.3 Safe Set (SS)

The SS is defined as all randomized subjects who have received at least one dose of the investigational drug and have safety records after drug administration.

For the safety analysis, all subjects should be analyzed based on the group of actually received study drug.

In this trial, baseline data are analyzed based on the FAS; primary efficacy endpoints are analyzed based on FAS (primary) and PPS; data of laboratory tests and AEs/Treatment-related AEs are analyzed based on the SS. In addition, ORR, DCR, and the percentage of subjects with SD will also be analyzed among subjects with evaluable efficacy.

Subjects with evaluable efficacy are defined as all subjects who have received at least 2 cycles (8 weeks) of apatinib treatment and received a tumor evaluation after 2 treatment cycles (8 weeks). If the efficacy reaches CR or PR, it should be confirmed 4 weeks later.

7 STATISTICAL ANALYSIS

7.1 General Principles

For continuous variables, the number of non-missing subjects, mean, standard deviation, median, minimum, and maximum should be listed. The decimal places of minimum and maximum are consistent with those of the records in the database. Mean and median retain one more decimal place than that of the source data recorded in the database; standard deviation retains two more decimal places than that of the source data recorded in the database. Unless otherwise stated, the "baseline" in this study is defined as the last non-missing measurement value obtained prior to the first use of the study drugs, including the measurements taken on the day of the first dose and prior to the administration of the first dose.

For categorical variables, a frequency table (frequency and percentage) will be used. Percentages retain one decimal place.

7.1.1 Significance level

For group comparisons, all hypothesis tests are performed using two-sided tests with an $\alpha = 0.05$. If P is ≤ 0.05, the difference of the test is considered to be statistically significant. A 95% confidence interval is used in this study.

7.1.2 Hypothesis test

The primary efficacy endpoint for this trial is OS.

The log-rank test is used for the comparison between the investigational drug group and control group:

Hypothesis:

H_0: The survival function of the apatinib group is the same as that of the placebo group;

H_1: The survival function of the apatinib group is different from that of the placebo group.

α level: 0.05 (two-sided).

7.1.3 Data principles

Missing dates: If a previous date is incomplete, which affects the calculation of subsequent dates, the previous date is estimated as follows:

If the date of first dose is missing, a randomization date will be replaced.

Apart from exceptional circumstances, the following filling rules apply to the missing dates of safety data events.

If the date of an event is completely missing, the date is not imputed. If the day is missing but the year and month of the event onset are the same as those of study treatment, then the missing day is imputed with the day of first date of study treatment, otherwise it is imputed with the first day of that month; if the day of the end date is missing, then it is imputed with the last day of that month.

If both month and day of an event onset date are missing but its year is the same as that of study treatment, then they are imputed with the month and day of starting study treatment, otherwise, they are imputed with 1 Jan.

If the date of death is completely missing, the last known survival date + 1 day of the subject is used as the date of death.

All imputed dates must be before the date of withdrawal of informed consent form, lost to follow-up, and death.

Missing data of laboratory tests, ECG, and vital signs are not imputed.

The laboratory test results are generally continuous numeric variables or character variables. If the continuous numeric variables contain special characters (such as < xx and > xx) in the eCRF, they are processed using the following rules:

(1) If a measured value is < xx or ≤ xx, half of the xx value will be used for analysis;

(2) If a measured value is > xx or ≥ xx, the value of xx will be used for analysis.

Dates in the date listings are listed as per CRF.

Descriptive statistics for summary include number of cases, number of missing cases, mean, standard deviation, median, minimum, and maximum.

The change from the baseline is defined as measured value in post-treatment follow-up - measured value at baseline.

The time (days) related to the efficacy endpoint: (post-treatment date - date of randomization) + 1. The selection of the post-treatment date follows the definition of efficacy endpoint.

The formula of OS calculation:

Death case: (date of death - date of randomization) + 1;

Surviving case: (date of the last survival - date of randomization) +1, as censored data.

The formula of PFS calculation:

The start and end dates of PFS for events with outcomes:

1: Date of radiographic PD (earliest date of progression of target lesions, non-target lesions, or new lesions) - date of randomization +1;

2: Date of death - date of randomization + 1.

If the above 1 and 2 are both met in the one case, the earliest time is taken.

The start and end dates of PFS for events without outcomes:

1: With imaging examination: date of the last adequate imaging examination - date of randomization + 1, as censored data;

2: Without imaging examination: 1 day (i.e., the date of censorship is the date of randomization), as censored data.

With a new anti-tumor treatment: date of the last imaging examination prior to or onset on starting the new anti-tumor treatment - date of randomization + 1, as censored data.

The start and end dates of PFS for those with an endpoint event but missing 2 or more (at least 18 weeks) imaging examinations:

1: Date of the last imaging examination before event observed - date of randomization +1, as censored data.

For the definition of new anti-tumor treatment, refer to Section 7.6.2.

7.2 Study Population

7.2.1 Subject Distribution

Description of the subject enrollment, dropout cases, completion of scheduled treatment, and premature discontinuation: number of subjects and percentage;

Distribution of subjects in each data set;

Distribution of subjects with evaluable efficacy;

Listing of dropout and premature discontinuation by subject: status of drug administration, reasons for trial termination, etc.

7.2.2 Protocol violations and deviations

Summarize and describe the protocol violations and deviations. See Appendix 1 for the classification of protocol violations.

7.2.3 Analysis of subject characteristics of two groups

Age: Describe using mean, standard deviation, maximum, minimum, and median.

Gender: Calculate the percentage.

Height and weight: Describe using mean, standard deviation, maximum, minimum, and median.

Tumor history: The course of disease is described using mean, standard deviation, maximum, minimum, and median. The percentages for pathological grade, clinical stage, presence of extrahepatic metastasis, presence of vascular invasion, alpha-fetoprotein (AFP < 400 μg/L, ≥ 400 μg/L; AFP < 200 μg/L, ≥ 200 μg/L), and cause of liver cancer (HBV infection: positive for either HbsAg or HbeAg; HCV infection: positive) are calculated.

Target lesion and non-target lesion at baseline: The sum of diameters of target lesion is described using mean, standard deviation, maximum, minimum, and median. The percentages for target lesion, location of target lesion, non-target lesion, and location of non-target lesion are calculated.

The percentages for history of primary lesion surgery, local ablation therapy, interventional procedure, radiotherapy, systemic treatment, past medical history, and drug allergy are calculated. Among them, the history of primary lesion surgery and past medical history are coded by the system organ class (SOC) and preferred term (PT) of the ICH Medical Dictionary for Regulatory Activities (MedDRA 21.0, Chinese version). The percentages for number of local ablation therapies and number of past interventional procedures are calculated.

The percentages for baseline ECOG PS, Child-Pugh classification of liver function, Barcelona clinic liver cancer (BCLC) staging, and hepatitis B and hepatitis C markers (including HbsAg, HbsAb, HbeAg, HbeAb, HbcAb, and HCV) are calculated.

Baseline vital signs (body temperature, heart rate, respiratory rate, and blood pressure): describe using mean, standard deviation, maximum, minimum, and median. The baseline is defined as the closest recent non-missing measured value to date of drug administration taken.

Baseline physical examination of various systems: The percentages of normal, abnormal, and unchecked results are calculated.

Baseline quality of life (QLQ-C30) scores and EORTC QLQ-HCC18 scores in various dimensions: describe using mean, standard deviation, maximum, minimum, and median.

7.3 Efficacy Evaluation

7.3.1 Analysis of primary efficacy endpoint

The primary efficacy endpoint for this study is OS. The OS analysis is performed based on FAS and PPS when at least 312 deaths are collected in this study (≥ 80% of patients died), of which FAS is the primary analysis set.

The survival functions of the apatinib group and placebo group are estimated by the Kaplan-Meier method, and the survival curve is plotted. The survival functions of the two groups are compared using the log-rank test with and without considering the stratification factors for randomization. In this study, the stratified log-rank test considering the stratification factors in randomization system is the primary test. In addition, under the assumption of proportional hazards, the Cox proportional hazards models with and without stratification factors are used to estimate the hazard ratio (HR) and calculate the corresponding 95% CI. In addition, the Cox proportional hazard model including other baseline factors may also be considered to estimate the HR between the two groups and calculate its 95% CI.

For the sensitivity analysis of OS, see Section 7.6.2.

7.3.2 Analysis of secondary efficacy endpoints

- Secondary efficacy endpoints of PFS and TTP are summarized using descriptive statistics based on FAS and PPS according to the above general principles and analyzed using similar methods as those for primary efficacy endpoint.
- Based on FAS and PPS, the 3-, 6-, and 12-month PFS rates, the Kaplan-Meier method was used to estimate rates and corresponding 95% CIs of the two groups are calculated using the log-log transformation according to the normal approximation with back transformation to CIs on the untransformed scale.
- The secondary efficacy endpoint DOR, including the investigator assessed results as per RECIST v1.1, is analyzed based on FAS and PPS. DoR will only be assessed in subjects whose BoR is PR or CR. The end date of tumor response must be consistent with the PD date of PFS or the date of death, and the Kaplan-Meier product-limit method will be used to calculate the median survival time of each treatment group and estimate the corresponding 95% CIs. The analysis is carried out using methods similar to those for the primary efficacy endpoint.
- Based on FAS, PPS, and evaluable patients, the ORR and DCR in the two groups are calculated and their 95% CIs are calculated using the Clopper-Pearson method. The difference of rates between groups are calculated and its 95% CIs are calculated using the Wald (normal approximation) method. The P values are calculated by the stratified Cochran–Mantel–Haenszel (CMH) (primary analysis) and the unstratified Pearson chi-square tests. Confirmed and unconfirmed CR/PR are analyzed, respectively.
- Based on FAS, PPS, and evaluable patients, the percentage of patients with SD for ≥ 6 weeks are calculated, and their 95% CIs are calculated using the Clopper-Pearson method. The difference between groups and its 95% CI are calculated using the Wald (normal approximation) method. The P values are calculated by the stratified CMH and the unstratified Pearson chi-square tests.

- Based on FAS and PPS, the 6- and 12-month OS rates, the Kaplan-Meier method is used to estimate rates and corresponding 95% CIs of the two groups are calculated using the log-log transformation according to the normal approximation with back transformation to CIs on the untransformed scale.
- Based on the FAS, the changes of the total quality of life scores from baseline between the two groups for each time point are compared by analysis of variance model (ANOVA). The changes from baseline in global health status, total quality of life score, and score of EORTC QLQ-HCC18 at each scale for each time point are compared by Wilcoxon rank-sum test between groups.

7.4 Safety Evaluation

7.4.1 Adverse events

AEs are encoded according to the current MedDRA 21.0 and then processed in statistical analysis.

The incidences of AEs, treatment-related adverse events (TRAEs), AEs leading to drug interruption, dose reduction, and drug discontinuation, and SAEs are summarized. If multiple AEs occurred in the same subject, the number of subject is counted as one when calculating the incidence; if a subject experiences multiple times of a same AE, the number of subject is counted as one when calculating the incidence of the AE.

AEs are summarized in the frequency table based on SOC and PT. The incidence is calculated based on SOC and symptom/sign (the number of cases is counted as number of subjects who have experienced at least one AE).

Severity of AEs and treatment-related adverse events: If a subject experiences multiple times of the same AE, the most severe one will be used in the analysis of the severity of the AE.

A list of cases of AEs and SAEs is provided.

The median time to the onset of the first event is calculated for the adverse events of special interest (AESIs). AESIs include hypertension, proteinuria, hand-and-foot syndrome, hemorrhage, electrocardiogram QT prolonged, and white blood cell count decreased. The specific classification definitions are shown in Appendix 2.

Time to the onset of first event (days) = onset date of first event − date of the first dose + 1

In addition, for the comparisons of the incidences of AEs between the treatment group and the control group, the Fisher's exact method may be used to calculate the P value, which is only used for clinical reference, rather than the deduction of statistical significance.

7.4.2 Drug exposure and dose modification

The drug exposure in the two groups is described using mean, standard deviation, maximum, minimum, and median.

The treatment interruption and dose modification during the trial are summarized.

The duration of drug exposure, actual dose intensity, and relative dose intensity of apatinib and placebo are summarized using mean, standard deviation, median, maximum, and minimum.

The specific definition rules of derived variables for the study drugs apatinib and placebo are shown in Appendix 3.

7.4.3 Vital signs

Measured values and their changes before and after treatment are described using mean ± standard deviation, maximum, minimum, and median.

7.4.4 Laboratory test parameters

Hematology, blood biochemistry, and AFP: Measured values and their changes before and after treatment are described using mean ±standard deviation, maximum, minimum, and median. Normal and abnormal changes after treatment are described in a shift table.

Urinalysis: Normal and abnormal changes after treatment are described using a shift table.

Stool routine: Normal and abnormal changes after treatment are described using a shift table.

The proportion of subjects with "abnormalities with clinical significance" will be presented, where the clinical significance of the abnormality is judged by the investigator.

7.4.5 ECG

ECG: The normal and abnormal changes before and after the treatment based on the normality and abnormality judged by the investigator are described.

7.4.6 Physical examination

The normality and abnormality before and after the treatment are described.

7.4.7 Concomitant medications

The usage and frequency of use of each drug for concomitant medications during the study period (including WBC-elevating drugs and platelet-elevating drugs used within 14 days before screening, any changes in the concomitant medications during the screening period, new concomitant medications after screening, medications for AEs, and concomitant medications during the follow-up period) are summarized.

7.5 Interim Analysis

None

7.6 Other Analyses

7.6.1 Subgroup analysis

Subgroup analyses are performed for the primary efficacy endpoint according to the following factors, and the forest plot on HR is produced:

Age (≤ 65 years old, > 65 years old)

ECOG PS (0, 1)

Presence of vascular invasion (yes, no)

Presence of extrahepatic metastasis (yes, no)

Presence of vascular invasion and/or extrahepatic metastasis (yes, no)

Treatment with sorafenib (yes, no)

Previous systemic treatment line(s) (one line, ≥ two lines)

HBV (positive, negative)

Alpha-fetoprotein (AFP < 400μg/L, ≥ 400μg/L)

Alpha-fetoprotein (AFP < 200 μg/L, ≥ 200 μg/L)

Child-Pugh score (Grade A, Grade B)

BCLC stage (B, C)

Previous liver transplant (yes, no)

7.6.2 Other Analyses

To assess the balance of new anti-tumor treatments between groups during the study, new anti-tumor treatments (including systemic treatment and local treatment) are summarized. In addition, the new anti-tumor treatments are classified, and the subgroup analyses of OS are performed. New anti-tumor treatments are classified into chemotherapy, targeted therapy, immunotherapy, and local treatments (TACE, surgery, ablation therapy, and radiotherapy). New anti-tumor treatments do not include traditional Chinese medicine and immunomodulators. Before database locking, classification of new anti-tumor treatments is finally determined and archived.

Other sensitivity analyses of OS include:

1) The OS of two groups will be compared by the Wilcoxon test and Max-Combo[1] method.

2) The survival time of subjects are censored at the start of new anti-tumor treatments.

3) In the OS analysis, some subjects who have received some other types of anti-tumor treatments (which may have a major impact on survival) during the study are excluded, such as subjects with subsequent PD1/PD-L1 treatments will be excluded.

4) Rank-preserving structural failure time model (RPSFTM) proposed by Robins and Tsiatis[2] : The OS of subjects who have received apatinib (or other targeted therapies) during the study anti-tumor treatment is adjusted using the RPSFTM method. This analysis will be based on the population defined by FAS and the sensitivity analysis method 2), respectively.

5) IPCW method proposed by Robins and Finkelstein[3]: Using the IPCW method to analyze the HR between the treatment group and control group, with reference to relevant medical literature, the baseline variables and time-related variables are planned to be considered in the IPCW model including ECOG PS, age, gender, presence of extrahepatic metastasis, number of line(s) of past treatment, BCLC grade, etc. If necessary, the duration of study drug exposure, and liver function parameters (such as ALT, AST, and bilirubin) may also be included. The bootstrap method is used to sample from the treatment group and control group, and the 95% CI of the HR is estimated.

6) The two-stage method proposed by Latimer et al.[4]: Use the two-stage method to adjust the OS of the subjects receiving the new anti-tumor treatment in the treatment group and control group, respectively. If necessary, relevant variables are included in the two-stage model (see the variables to be included in the IPCW method).

In addition, the stratification factors for randomization in the OS analysis in Section 7.3 will be based on the stratification information in the subject randomization system unless otherwise specified. If the stratification factor information in the randomization system and EDC clinical database are inconsistent (the stratification information in the randomization system is incorrect) for some subjects, their stratification information in the EDC clinical database will be used for the sensitivity analysis.

8 STATISTICAL ANALYSIS SOFTWARE

Jiangsu Hengrui Pharmaceuticals Co., Ltd. is responsible for the data management, and the Department of Biostatistics, School of Public Health, Nanjing Medical University undertakes the statistical analysis of this study. All statistical analyses will be performed using SAS 9.4.

9 REFERENCES

[1]. Public Workshop: Oncology Clinical Trials in the presence of non-proportional hazards. 2018. https://healthpolicy.duke.edu/sites/default/files/atoms/files/oncology_trials_workshop_meeting_summary_0.pdf.

[2]. Robins, J. and Tsiatis, A. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Communications in Statistics – Theory and Methods 1991; 20(8):2609-2631.

[3]. 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.

[4]. 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.

10 ATTACHED TABLES

Appendix 1. Classification table of protocol deviations

Class	Sub-Class	Description of Protocol Deviation	Grade
1-Inclusion/Exclusion Criteria	Eligibility	Not meet a certain inclusion criterion, and/or not meet a certain exclusion criterion	Critical
1-Inclusion/Exclusion Criteria	Eligibility	Misunderstanding of the inclusion/exclusion criteria causes the subject enrollment not meeting the protocol requirements	Critical
2-Informed Consent	Informed consent criteria	The subject has been randomized or started treatment with the study drug before signing the informed consent form	Critical
2-Informed Consent	Informed consent criteria	The subject signs the informed consent form after starting any study-related procedures	Critical
2-Informed Consent	Informed consent criteria	The subject signs an incorrect version of the informed consent form (the latest version is not signed)	Critical
2-Informed Consent	Informed consent criteria	The contact information of the ethics committee or study site is not shown or recorded in the informed consent form	Major
2-Informed Consent	Informed consent criteria	The informed consent form is not dated (including by the subject and/or the investigator)	Major
2-Informed Consent	Informed consent criteria	The informed consent process is not recorded or the record is incomplete	Major
2-Informed Consent	Informed consent criteria	The informed consent procedure is not followed	Critical
2-Informed Consent	Informed consent criteria	The investigator who signs the informed consent form is not in the authorization form or is not authorized	Major
3-Visit Schedule	Criteria for visit schedule	Items scheduled for the visits of the treatment period are not carried out: physical examination, vital sign measurement, quality of life scoring, and ECOG scoring	Minor
3-Visit Schedule	Criteria for visit schedule	Items scheduled for the visits of the treatment period are not carried out: except for physical examination and/or vital sign measurement, quality of life scoring, and ECOG scoring	Major
3-Visit Schedule	Criteria for visit schedule	The first imaging examination after treatment and assessed as PD is carried out beyond the time window	Major
3-Visit Schedule	Criteria for visit schedule	The image examinations other than the ones described above are carried out beyond the time window, or the number of image examinations is more than required	Minor
3-Visit Schedule	Criteria for visit schedule	The imaging examination is not done, or the examination method or technique is inconsistent with that at baseline	Critical
3-Visit Schedule	Criteria for visit schedule	The end-of-treatment visit/safety visit is carried out beyond the time window	Minor
3-Visit Schedule	Criteria for visit schedule	The end-of-treatment visit/safety visit is not done, or there are missing examination items	Major
3-Visit Schedule	Criteria for visit schedule	The survival follow-up is carried out beyond the time window/is missing	Minor

Class	Sub-Class	Description of Protocol Deviation	Grade
3-Visit Schedule	Criteria for visit schedule	Other than those mentioned above, the subject has not conducted the visit as required by the protocol or has not completed the examination and other procedures required for the visit	Depends on the situation (minor or major)
4-Subject Diary	Criteria for study procedures	Less than 80% of the diary card is completed	Minor
4-Subject Diary	Criteria for study procedures	The subject has not brought the diary card to the study site	Major
5-Concomitant Medications	Criteria for concomitant medications	The subject has used a drug prohibited by the trial protocol	Depends on the situation (major or critical)
6-Subject Discontinuation	Criteria for study procedures	The subject meets the criteria for study withdrawal or discontinuation, but the subject has not discontinued the study	Critical
6-Subject Discontinuation	Criteria for study procedures	Discontinues the study drug but does not follow the protocol requirements	Major
7-Study Drugs	Criteria for study procedures	No drug is administered within 48 h after randomization	Minor
7-Study Drugs	Criteria for study procedures	Cases that the storage conditions of apatinib exceed the requirements for a short period of time that does not affect the quality of the drug according to the requirements on the instructions for the storage and use of apatinib	Major
7-Study Drugs	Criteria for study procedures	Cases that the sponsor shall be contacted immediately according to the requirements on the instructions for the storage and use of apatinib	Critical
7-Study Drugs	Compliance to study drugs	Missed dose or overdose of the study drug	Depends on the situation (major or critical)
7-Study Drugs	Compliance to study drugs	The total duration of drug interruption during each treatment cycle exceeds 2 weeks	Critical
7-Study Drugs	Compliance to study drugs	The subject is given a drug to be used by other group	Critical
7-Study Drugs	Criteria for study procedures	The study site or the subject has discarded the bottle of used drug Discrepancies in the inventory of the study drugs	Minor
7-Study Drugs	Other criteria	No continuous temperature records	Major
7-Study Drugs	Compliance to study drugs	In addition to the above, other non-compliance with the use of the study drugs	Depends on the situation (minor or major)
8-Safety Monitoring	SAE	The report of SAEs has not met the time limit required by GCP or related regulations	Critical
8-Safety Monitoring	Review of safety reports	The investigator has not reviewed and signed the safety reports of subjects in time	Major

Class	Sub-Class	Description of Protocol Deviation	Grade
8-Safety Monitoring	AESIs	Incorrect and inappropriate reporting time of AEs	Critical
9-Study Procedures	Management criteria	The responsibility authorization form has not been updated before the start of the study procedures	Minor
9-Study Procedures	Criteria for study procedures	The study operation of the investigator has exceeded the scope of authorization	Major
9-Study Procedures	Management criteria	Non-compliance with other study-related procedures in addition to the above-mentioned clinical trial protocol and drug manual.	Depends on the situation
9-Study Procedures	Criteria of ethical approval	Use study-related documents that are not ethically approved	Critical
9-Study Procedures	Monitoring of entire study by principal investigator	The motoring is insufficient, especially in terms of the lack of evidence for the monitoring by the principal investigator regarding the safety of subjects	Major
9-Study Procedures	Criteria for source documents	The study site has changed the placement location of the source records of subjects without notifying the clinical research associate (CRA) or providing the CRA with access rights; or the source records are stored at the study site without giving the CRA sufficient rights to access the source records (including HIS/LIS) (per ICH GCP R2 4.1.4, the investigator/study site shall allow the sponsor's monitoring agency to conduct monitoring and verification.)	Major
9-Study Procedures	Randomization procedures	Error in information of randomized stratification	Major

Appendix 2. Adverse events of special interest

Adverse Event of Special Interest	Preferred Terms of AE
Hypertension	Blood pressure increased
Hypertension	Hypertension
Proteinuria	Protein urine present
Proteinuria	Urine protein
Proteinuria	Proteinuria
Hand-and-Foot Syndrome	Palmar-plantar erythrodysesthesia syndrome
Hemorrhage	Upper gastrointestinal hemorrhage
Hemorrhage	Gastrointestinal hemorrhage
Hemorrhage	Gingival bleeding
Hemorrhage	Anal hemorrhage
Hemorrhage	Hematochezia
Hemorrhage	Melena
Hemorrhage	Mouth hemorrhage

Adverse Event of Special Interest	**Preferred Terms of AE**
Hemorrhage	Mucosal hemorrhage
Hemorrhage	Subcutaneous hemorrhage
Hemorrhage	Purpura
Hemorrhage	Cerebral hemorrhage
Hemorrhage	Brain stem hemorrhage
Hemorrhage	Pharyngeal hemorrhage
Hemorrhage	Bronchial hemorrhage
Hemorrhage	Hemoptysis
Hemorrhage	Epistaxis
Hemorrhage	Hepatic hemorrhage
Hemorrhage	Scleral hemorrhage
Hemorrhage	Vaginal hemorrhage
Electrocardiogram QT prolonged	QT interval prolonged
White blood cell count decreased	White blood cell count decreased
White blood cell count decreased	Leukopenia

Appendix 3. Derivative variables for apatinib/placebo

Variable	**Apatinib/Placebo**
Protocol-Specified Method of Administration	Once a day, 750 mg each time, every 4 weeks in a treatment cycle
Duration of Drug Exposure (days)	Date of last administration - date of first administration + 1
Planned Duration of Treatment (days)	Date of last administration - date of first administration + 1
Actual Cumulative Doses of All Cycles (mg)	Summarize the actual dose (mg) in all treatment cycles: [duration of drug exposure (day) - (end date of dose modification - start date of dose modification) (day)] ×750 (mg) + (end date of dose modif ication - start date of dose modification) (day) ×dose after modification (mg) - (end date of dose interruption - start date of drug interruption) (day) ×interrupted dose (mg)
Planned Dose Intensity (mg/day)	750 (mg/day)
Actual Dose Intensity (mg/day)	Actual cumulative dose in all treatment cycles (mg) / [planned duration of treatment (day)]
Relative Dose Intensity (%)	100 ×[actual dose intensity (mg/day)] / [planned dose intensity (mg/day)]