Official Title: Phase 3, Open-Label, Randomized Study of Futibatinib Versus Gemcitabine Cisplatin Chemotherapy as First-Line Treatment of Patients with Advance Cholangiocarcinoma Harboring FGFR2 Gene Rearrangements

NCT Number: NCT04093362

Document Date: SAP Version 1.0: 26 March 2021

STATISTICAL ANALYSIS PLAN

VERSION: 1.0

DATE OF PLAN: 26 MAR 2021

STUDY DRUG: Futibatinib (TAS-120)

PROTOCOL NUMBER: TAS-120-301

IND Number: 121062

STUDY TITLE:

A Phase 3, Open-Label, Randomized Study of Futibatinib Versus Gemcitabine-Cisplatin Chemotherapy as First-Line Treatment of Patients with Advanced Cholangiocarcinoma Harboring *FGFR2* Gene Rearrangements

FOENIX-CCA3

SPONSOR:

Taiho Oncology, Inc.

101 Carnegie Center

Princeton, NJ USA 08540

+1 (609) 750-5300

This trial was conducted in accordance with the ethical principles of Good Clinical Practice (GCP), according to the International Conference on Harmonization (ICH) Harmonized Tripartite Guideline.

SIGNATURE PAGE

A Phase 3, Open-Label, Randomized Study of Futibatinib Versus Gemcitabine-Cisplatin Chemotherapy as First-Line Treatment of Patients with Advanced Cholangiocarcinoma Harboring *FGFR2* Gene Rearrangements

FOENIX-CCA3

Protocol TAS-120-301

Prepared by:

[redacted] PhD
[redacted] Biostatistics

[redacted]
Electronically signed by: [redacted]
Reason: I am the author.
Date: Jul 27, 2021 12:17 EDT
Signature

Date

Approved by:

[redacted] MD, PhD
[redacted]
Biostatistics

[redacted]
Electronically signed by: [redacted]
Reason: I have read and approve
Date: Jul 27, 2021 12:48 EDT
Signature

Date

[redacted] MD
[redacted] Clinical Development

[redacted]
Electronically signed by: [redacted]
Reason: I have read and approve
Date: Jul 28, 2021 17:07 GMT+2
Signature

Date

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Term
AE	adverse event
ALT	alanine aminotransferase
ANC	absolute neutrophil count
ARP	All randomized population
ASCO	American Society of Clinical Oncology
AST	aspartate aminotransferase
BCRP	breast cancer resistance protein
CCA	cholangiocarcinoma
CI	confidence interval
CMH	Cochran-Mantel-Haenszel
CR	complete response
CSR	clinical study report
CT	computed tomography
ctDNA	circulating tumor DNA
CYP	Cytochrome P450
DCR	disease control rate
DMC	Data Monitoring Committee
DOR	duration of response
DRAE	Drug-related adverse event
eCCA	extrahepatic cholangiocarcinoma
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	electronic case report form
EDC	electronic data capture
EOT	end of treatment
FDA	Food and Drug Administration
FGF	fibroblast growth factor
FGFR	fibroblast growth factor receptor
GCP	Good Clinical Practice
HR	Hazard Ratio
I.V.	intravenous
iCCA	intrahepatic cholangiocarcinoma
ICF	informed consent form
ICH	International Conference on Harmonisation
ICR	independent central review
IEC	independent ethics committee
IRB	institutional review board
IXRS	Interactive Voice/Web Response System

KM	Kaplan-Meier
MRI	magnetic resonance imaging
NCI-CTCAE	National Cancer Institute – Common Terminology Criteria for Adverse Events
ORR	objective response rate
OS	overall survival
PD	progressive disease
PFS	progression-free survival
PFS2	PFS on next-line therapy
P-gp	P-glycoprotein
PI	prescribing information
PK	pharmacokinetic(s)
Pop PK	population pharmacokinetics
PP	Per Protocol
PR	partial response
PRO	patient-reported outcomes
PS	performance status
QD	daily
QTcF	Fridericia's corrected QT interval
RECIST 1.1	Response Evaluation Criteria in Solid Tumors, Version 1.1
SAE	serious adverse event
SAP	Statistical Analysis Plan
SD	stable disease
SOP	standard operating procedure
TEAE	treatment-emergent adverse event
TID	three times per day
TOI	Taiho Oncology, Inc.
ULN	upper limit of normal
TRAEs	Treatment-related AEs (TRAEs)
WBC	white blood cell
WOCBP	woman/women of child-bearing potential

1. INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to describe the planned analyses and data displays to be included in the Clinical Study Report (CSR) for Protocol TAS-120-301. The SAP is developed according to the Protocol TAS-120-301, dated 13 September 2019.

Genomic aberrations in *FGF* or the *FGFR* are a reported genetic modification in CCA. In iCCA, *FGFR2* gene rearrangements, including gene fusions, have been identified as an early driver of oncogenic events. Futibatinib, a selective small molecule FGFR inhibitor, has shown promising antitumor activity in preclinical studies and ongoing clinical studies against a variety of tumor types harboring *FGFR* aberrations.

The Phase 3 study will evaluate the safety and efficacy of futibatinib to compare with the current standard of care (gemcitabine-cisplatin chemotherapy) in the first-line treatment of patients with advanced CCA harboring *FGFR2* gene rearrangements (including gene fusions).

Study measurements and assessments, planned statistical methods, and derived variables are summarized in this plan. Planned tables, figures, and listings are specified. All decisions regarding final analyses, as defined in this SAP document, will be made prior to locking the database. Any deviations from these guidelines will be documented in the CSR.

2. STUDY OBJECTIVES AND ENDPOINTS

2.1. Study Objectives

2.1.1. Primary Objective

The primary objective of the study is to compare futibatinib (TAS-120) monotherapy relative to the current standard of care (gemcitabine-cisplatin chemotherapy) in the first-line treatment of patients with advanced iCCA harboring *FGFR2* gene rearrangements (including gene fusions) with respect to progression free survival as assessed by independent review committee.

2.1.2. Secondary Objectives

The secondary objectives of this study are:

- To compare the objective response rate (ORR) of patients treated with futibatinib versus current standard of care (gemcitabine-cisplatin chemotherapy);
- To compare the disease control rate (DCR) of patients treated with futibatinib versus current standard of care (gemcitabine-cisplatin chemotherapy);
- To compare overall survival (OS) of patients treated with futibatinib versus current standard of care (gemcitabine-cisplatin chemotherapy);
- To compare PFS per Investigator assessment, of patients treated with futibatinib versus current standard of care (gemcitabine-cisplatin chemotherapy);
- To compare the safety and tolerability of treatment with futibatinib versus current standard of care (gemcitabine-cisplatin chemotherapy);

2.1.3. Exploratory Objectives

- To compare PFS on next-line therapy (PFS2) of patients treated with futibatinib versus current standard of care (gemcitabine-cisplatin chemotherapy);
- To compare duration of response (DOR) of patients treated with futibatinib versus current standard of care (gemcitabine-cisplatin chemotherapy);
- To compare patient-reported outcomes (PRO) of patients treated with futibatinib versus current standard of care (gemcitabine-cisplatin chemotherapy);
- To assess the population pharmacokinetics (Pop PK) of futibatinib and to explore the relationship between PK and efficacy or toxicity.

2.2. Study Endpoints

2.2.1. Primary Endpoint

The primary endpoint of the study is progression-free survival (PFS) as assessed by independent central review (ICR), defined as the time from the date of randomization to the date of the first documentation of disease progression using RECIST 1.1 criteria or death due to any cause. PFS censoring rules will follow the FDA guidance of 2007 (FDA 2007).

2.2.2. Secondary Endpoints

The secondary efficacy endpoints are:

- ORR, defined as the proportion of patients who have best overall response of CR or PR as determined by ICR using RECIST 1.1.
- DCR**,** defined as the proportion of patients experiencing a best overall response of stable disease (SD), PR, or CR (per RECIST 1.1), based on ICR.
- OS, defined as the time from the date of randomization to the date of death from any cause. Patients who are lost to follow-up and those who are alive at the date of data cut-off will be censored at the date the subject was last known alive, or date of data cut-off.
- PFS per Investigator assessment, defined as the time from date of randomization to the date of disease progression based on Investigator assessment of radiographic images or death, whichever occurs first.

The secondary safety and tolerability endpoints are:

- Treatment-emergent adverse events (TEAEs), including serious adverse events (SAEs), clinical laboratory tests, vital signs, ophthalmological exams, and 12-lead electrocardiogram (ECG).

2.2.3. Exploratory Endpoints

- DOR, defined as the time from the first documentation of response (per RECIST 1.1 and based on central assessment of radiologic images) to the first documentation of objective tumor progression or death due to any cause, whichever occurs first.
- PRO, the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30) and the Cholangiocarcinoma and Gallbladder Cancer module (EORTC QLQ-BIL21).
- PFS on next-line therapy (PFS2), defined as the time from randomization to subsequent disease progression per ICR after initiation of next-line therapy or death from any cause (whichever occurs first).
- Population pharmacokinetics (Pop PK) of futibatinib.

3. STUDY DESCRIPTION

3.1. Study Design

Study TAS-120-301 is an open-label, multinational, parallel two-arm, randomized Phase 3 study evaluating the efficacy and safety of futibatinib versus gemcitabine-cisplatin chemotherapy as first-line treatment in patients with advanced, metastatic, or recurrent unresectable iCCA harboring *FGFR2* gene rearrangements. Approximately 216 patients will be centrally randomized in a 1:1 ratio to receive futibatinib or gemcitabine-cisplatin chemotherapy (control arm).

Patients in the experimental arm may continue to receive futibatinib until documentation of progressive disease (PD) per RECIST 1.1, or until other withdrawal criteria are met, whichever comes first. However, treatment may continue following PD per RECIST 1.1 if the patient is clinically stable and is considered by the Investigator to be deriving continued clinical benefit from futibatinib.

Patients in the control arm may receive gemcitabine-cisplatin chemotherapy for up to 8 cycles or until PD or other withdrawal criteria are met, whichever comes first. Patients who discontinue gemcitabine-cisplatin due to documented disease progression (by ICR) may receive treatment with futibatinib ("crossover"), if medically appropriate in the opinion of the Investigator and if criteria for futibatinib treatment are met. Patients who discontinue gemcitabine-cisplatin for other reasons (eg, safety, withdrawal of consent, Investigator decision) are eligible for crossover following subsequent ICR-confirmed PD.

The study design is summarized in Figure 1.

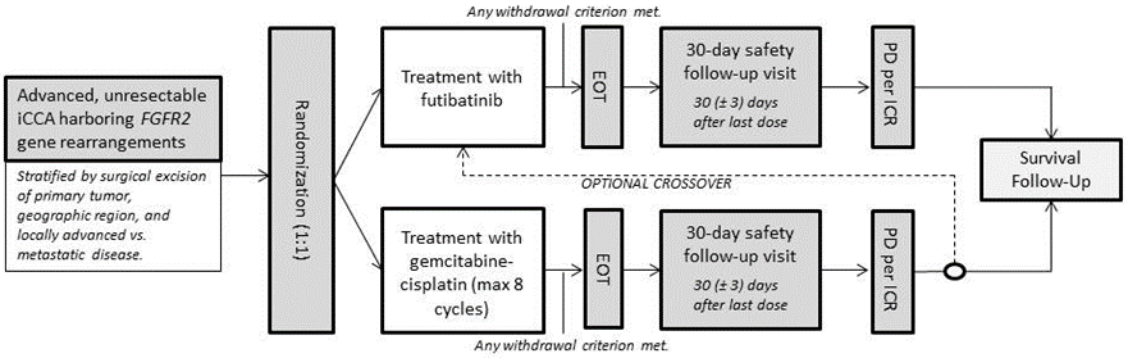

Figure 1: Study Schema

3.2. Randomization and Treatment Assignment

Randomization will occur centrally using an Interactive Voice/Web Response System (IXRS). Patients will be assigned randomly in a 1:1 ratio to the treatment arms as follows:

a. Experimental Arm: futibatinib (20 mg QD)
b. Control Arm: gemcitabine-cisplatin (I.V. infusion)

- Cisplatin 25 mg/m^2 in 1000 mL 0.9% saline by intravenous (I.V.) infusion over 1 hour, followed by 500 mL 0.9% saline over 30 minutes; and

- Gemcitabine 1000 mg/m^2 in 250-500 mL 0.9% saline by I.V. infusion over 30 minutes, beginning after completion of the cisplatin and saline infusions.

Randomization will be stratified according to the following factors:

- Prior surgical excision of the primary tumor (yes vs. no).
- Geographic region (Asia vs. rest of the world).
- Locally advanced vs. metastatic disease.

3.3. Determination of Sample Size

Calculations are done with EAST 6.4 software.

Assuming median PFS of 10 months in the Experimental Arm and 6 months in the Control Arm, a total of 162 PFS events will be required for the primary analysis to detect a hazard ratio of 0.60 with 90% power using a log-rank test at a two-sided 5% level of significance.

Based on the assumption that enrolment will continue for about 32 months, and that about 10%/year of the patients will drop out, a total of approximately 216 patients randomized in a 1:1 ratio will be needed to observe at least 162 PFS events approximately 8 months after the last patient randomization.

4. STUDY PERIODS, TREATMENT REGIMENS, AND POPULATIONS FOR ANALYSIS

4.1. Study Periods for Analyses

For each patient, the Study Duration is defined as the time from day of ICF signature to the last day of Disease Assessment Follow-up/Survival Follow-up.

Study periods are defined in Table 1.

Table 1: Definition of Study Periods for Analysis

Period	Definition
Baseline	Up to 3 days before the date of the first dose of study drug.
On-Treatment	From the date of the first dose of study drug to 30 days after the last dose of study drug. Unless otherwise specified, the on-treatment period will be the basis for the summaries of safety.
Safety Follow-Up	From the date of last dose of study drug through 30 days (±3 days) following the last dose of study drug or until the start of new anticancer therapy, whichever comes first.
Survival Follow-up	From disease progression or new anticancer therapy, whichever occurs first, survival follow-up every 12 weeks (±2 weeks) until withdrawal of consent, death, or loss to follow up, or study completion.
Study Completion	The study will be considered complete 12 months after the last event has been observed for the primary analysis of efficacy or the trial is halted early for any reason.

Abbreviations: ICF=informed consent form.

4.2. Populations for Analysis

- **All Enrolled Population:** All patients who signed an informed consent form and were registered into the IVRS.
- **All Randomized Population** (AR): All patients who were randomized to any treatment group in the study. This is the primary population for analyses of demography, protocol deviations, baseline characteristics and efficacy.
- **All Treated Population** (AT) : All patients who received at least one dose of study drug. This is the primary population for dosing and safety analyses.
- **Per-Protocol Population** (PP): All randomized patients who have no relevant protocol deviations. Relevant protocol deviation will be identified prior to unblinding of the study. For patients who have a relevant deviation during the study, data collected before the point of deviation can be included in the analysis preformed for this population.

- **Pharmacodynamic / Biomarker Evaluable Set**: All patients in the All Treated Population who have evaluable pharmacodynamic/biomarker data for analyses.
- **Pharmacokinetic Evaluable Set**: All patients who received futibatinib and have evaluable plasma data for analysis

4.3. Timing of Analysis

The primary analysis will be performed when approximately 162 PFS events are accumulated. There is no planned interim analysis for efficacy or futility. Interim analyses for safety monitoring will occur when approximately 25%, 50%, and 75% of the planned total number of patients have been treated. A Data Monitoring Committee (DMC) will meet to assess the cumulative safety data and recommend study continuation, discontinuation, or study modification. A description of the roles and responsibilities of the DMC and details of the review processes are provided in a separate DMC charter.

5. STATISTICAL METHODOLOGY

5.1. General Methods

All recorded data will be presented in listings.

Unless otherwise stated, descriptive statistics and summaries will be presented by treatment arm and overall using All Randomized Population (ARP).

Descriptive statistics will be provided for selective demographics, safety, PD and biomarker data. Descriptive statistics on continuous data will include non-missing observations (n), means, medians, standard deviations and ranges, while categorical data will be summarized using frequency count and percentage. Graphical summaries of the data may also be provided.

Confidence interval (CI) for proportions will be estimated using an exact method proposed by Clopper-Pearson (Clopper and Pearson, 1934).

Time to event endpoints will be estimated using the Kaplan–Meier method. The number of events and censorings will be reported. When appropriate, the median along with the corresponding log-log transformed 95% CI will be estimated. Survival rates at fixed time points (eg, PFS at 6 months) will be derived from the Kaplan–Meier estimate and the corresponding CI will be derived based on the Greenwood formula (Kalbfleisch and Prentice, 1980).

All the analyses for this study will be performed using SAS®, Version 9.3 or a higher version.

5.2. Data Quality

The sponsor will ensure that the clinal trial is conducted with a quality management system that will define quality tolerance limits in order to ensure human subject protection and reliability of study results. Data will be generated, documented, and reported in compliance with the protocol, ICH GCP, and applicable regulatory requirements.

Although this trial is being conducted as an open-label study, analyses or summaries generated by randomized treatment assignments, actual treatment received, and biomarker status will be limited and documented.

5.3. Disposition and Baseline Characteristics

5.3.1. Patient Disposition

The total number of patients enrolled by treatment group will be summarized along with the reason for not being randomized. The table will include the number that withdrew prior to completing the study, and the primary reasons for withdrawal. The number of patients completing treatment, and the primary reasons for treatment discontinuation will also be summarized.

A list of all randomized patients will be provided showing each patients study/treatment completion information, including first and last dosing date, off study date and reason for going off-study, the reason for premature study withdrawal/treatment discontinuation, if applicable, will be presented. A listing for patients not randomized will also be provided, showing each patient's race, gender, age, consent date and reason for not being included.

A listing of analysis populations will be provided.

5.3.2. Protocol Deviations

Important protocol deviation (ICH E3 Q&A (R1)) is defined as a subset of protocol deviations that may significantly impact the completeness, accuracy, and/or reliability of the study data or that may significantly affect the patient's rights, safety or well-being. These include important informed consent form (ICF) issues, inclusion/exclusion criteria not met, withdrawal criteria not followed; wrong treatment, incorrect dose/overdose based on protocol definitions, important deviations based on protocol design, and other important Good Clinical Practice (GCP) deviations.

Important protocol deviations will be summarized and listed in the Clinical Study Report (CSR).

According to US Food and Drug Administration (FDA) guidance on conduct of clinical trials of medical products during COVID-19 public health emergency, updated on March 2020 and updated on January 27, 2021, protocol deviations related to COVID-19 will be listed in the CSR.

5.3.3. Demographic and Other Baseline Characteristics

All demographic and other baseline characteristic summary statistics and analyses are based on the latest one collected prior to the first dose of study drug or date of randomization for untreated patients. Baseline characteristics data including age, sex, race, ethnicity, height (cm), weight (kg) and baseline ECOG performance status will be listed individually by patient and summarized overall and by treatment arm.

Disease characteristics at initial diagnosis (primary tumor location, histology type, tumor grade, TNM classification, AJCC staging), Extent of disease at baseline (locally advanced or metastatic) will be summarized and listed.

5.3.4. Medical History

Medical history will be coded according to the Medical Dictionary for Regulatory Activities (MedDRA). The actual version of the MedDRA coding dictionary will be noted in the statically tables and clinical study report. The number and percentage of patients in each medical history category (by MedDRA system organ class and preferred term) will be summarized overall and by treatment arm. The system organ class (SOC) will be presented in alphabetical order, and the preferred terms will be presented in alphabetical order within each SOC. Patients reporting more than one condition/diagnosis will be counted only once in each row (SOC or preferred term).

General medical history will be listed by patient, and pre-treatment events will be tabulated.

5.3.5. Prior/Concomitant Therapy

A medication is considered prior if the medication stop date is before treatment start date. Concomitant therapy includes all medications/therapies administered from the time ICF is signed through 30 days after the last dose of TAS0120 or until the start of new anticancer therapy.

Prior/concomitant non-study medication will be classified by anatomic and therapeutic classes. Agents and medication will be reported using the generic name. A listing by patient will also be provided. Prior medications will be flagged in the listing. Medication terms will be coded using WHO drug dictionary (WHODDE B2).

5.4. Efficacy Analyses

Unless otherwise specified, all efficacy analyses will be based on All Randomized Population.

5.4.1. Primary Efficacy Analyses

The primary efficacy endpoint of this study is ICR-assessed PFS, which is defined as the time from randomization to the first occurrence of progression as assessed by ICR or death from any cause, whichever comes first. The primary analysis of the study will test the equality of PFS distribution in the futibatinib (experimental arm) and in the gemcitabine-cisplatin (GC, control arm), as follows:

$$H_0: PFS_{futibatinib} = PFS_{GC} \text{ vs. } H_1: PFS_{futibatinib} \neq PFS_{GC}$$

The treatment comparison will be performed using a two-sided log-rank test at 0.05 significant level stratified by prior surgical excision of the primary tumor (yes or no), geographic region (Asia or rest of the world) and locally advanced or metastatic disease. If the null hypothesis is rejected and the observed hazard ratio (HR) is favorable for the fubibatinib arm, it is showed that fubibatinib treatment arm has statistically significant longer PFS than that of the GC arm.

The HR and the corresponding 95% CIs will be estimated using the Cox regression model stratified by the same stratification factors. The median PFS, and the PFS rates at 6, 12 and 18 months, will be calculated using the KM product-limit estimates for each treatment arm, and presented with two-sided 95% CIs. The Kaplan-Meier plot will be used to provide a visual description of the differences in duration of PFS across treatment arms. The investigator-assessed PFS will also be analyzed using the same models. Agreement/Disagreement between investigator assessment and assessment by the ICR will be summarized.

The PFS censoring rules in this SAP and definition of progression date follow the Food and Drug Administration (FDA) "Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics (2007)". The detail event definition and censoring rule are described in Table 2.

Table 2: Censoring Rules for Progression-Free Survival

No.	Situation	End Date	Outcome
1	Documented PD between scheduled visits	Date of the first assessment that determined PD	PFS event
2	Death during the study with no prior PD	Date of death	PFS event
3	Patients still on treatment without PD as of data cut-off [a]	Date of the last tumor assessment [b]	Censored
4	Treatment discontinuation for reasons other than PD or death, and no post baseline tumor assessments	Date of the first dose	Censored
5	Treatment discontinuation for reasons other than PD or death with post baseline tumor assessments	Date of the last tumor assessment	Censored
6	New anticancer treatment started	Date of the last tumor assessment before start of new treatment	Censored

No.	Situation	End Date	Outcome
7	Death or PD after two or more missed tumor assessments [c]	Date of the last tumor assessment before missed assessments	Censored
8	No baseline or unreadable baseline assessment but readable post baseline assessments	Date of the first dose	Censored

[a.] This refers to patients who were still receiving study treatment at time of data cutoff.
[b.] For PFS analysis, the date of last tumor assessment refers to the date of last adequate tumor assessment.
[c.] Two or more missed tumor assessments is defined as having either one of the following two durations being longer than for the first 12 weeks, or thereafter 3:
Duration between two consecutive tumor assessments
Duration between the last tumor assessment and death or PD

The following sensitivity analyses will be performed.

- PFS will be compared between treatment groups using unstratified log-rank test.
- To investigate impact of informative censoring on study outcomes, PFS will be analyzed using different censoring rules, such as:
    - Counting patients who lost to follow-up and those who started new anticancer therapy as events.
    - Using the actual reported date of progression by IRC or death to define PFS regardless of missing assessments, treatment discontinuation, or use of new anti-cancer therapy.
- PFS will be compared using the primary censoring rules based on per-protocol population.

5.4.2. Secondary Efficacy Analyses

5.4.2.1. Overall Response Rate (ORR)

ORR will be calculated for each treatment arm with exact 95% CIs using the method of Clopper and Pearson. The analysis of ORR for comparison between the two arms will be performed at the two-sided significant level of 0.05 by using a stratified Cochran-Mantel-Haenszel (CMH) test with the stratification factors in it.

5.4.2.2. Disease Control Rate (DCR)

DCR will be calculated for each treatment arm with exact 95% CIs using the method of Clopper and Pearson. The analysis of DCR for comparison between treatment arms will be performed using the CMH test stratified by the stratification factors used in the randomization.

5.4.2.3. Overall Survival (OS)

Time to event endpoint OS will be analyzed using the same log-rank tests as described in the primary analysis of PFS. The detail event definition and censoring rule of OS are described in Table 3.

Table 3: Censoring Rules for Overall Survival

Situation	End Date	Outcome
Death before cut-off	Date of death	Death
Death after data cut-off	Date of data cut-off	Censored event
Patient still alive at data cut-off	Date of data cut-off	Censored event
Patient discontinued treatment due to any reason before data cut-off	Date last known to be alive	Censored event

5.4.2.4. PFS per Investigator assessment

PFS per investigator assessment is defined as the time from date of randomization to the date of disease progression based on Investigator assessment of radiographic images or death, whichever occurs first. The similar approach for analysis of primary endpoint PFS per ICR descried in 5.4.1 will be used.

5.4.3. Exploratory Efficacy Analyses

5.4.3.1. Duration of response (DOR)

DOR is defined as the time from first documented response (CR or PR), as assessed by the investigator or local radiologist, to the date of first documented disease progression or death due to underlying cancer. Analysis of the DOR will incorporate data only from the subset of patients in both treatment arms that achieved a CR/PR. Censoring rules for DOR are summarized in Table 4.

Table 4: Censoring Rules for Duration of Response

Situation	Date of Progression or Censoring	Outcome
No progression nor death, no new anticancer therapy initiated	Last adequate disease assessment	Censor (non-event)
No progression nor death, new anti-cancer therapy initiated	Last adequate disease assessment before new anti-cancer therapy initiated	Censor (non-event)
Death or progression after ≥ 2 missed adequate disease assessments	Last adequate disease assessment prior to the after ≥ 2 missed adequate disease assessments	Censor (non-event)
Death or progression after ≤ 1 missed adequate disease assessments	PD or death	End of response (Event)

DOR (months) = ([Date of progression/death due to underlying cancer – date of first response] + 1) / 30.5

As there is no expectation for treatment arm balance, analyses of DOR will not incorporate stratification factors and will not produce a p-value but only be summarize the treatment arm estimates and confidence intervals.

5.4.3.2. Patient Reported Outcomes (PRO)

The following PRO assessments will be performed as exploratory analyses:

- European Organization for Research in the Treatment of Cancer – Quality of Life Questionnaire (EORTC-QLQ-C30)
- The Cholangiocarcinoma and Gallbladder Cancer module (EORTC QLQ-BIL21).

The PRO instruments will be completed by patients using paper-based questionnaires. All assessments should be completed without assistance from anyone. The PRO questionnaires will be administered on the days specified in the Schedule of Events.

Patients who discontinued treatment due to toxicity should continue the PRO assessments. PRO data collection will be discontinued due to patient withdrawal of consent, initiation of subsequent anti-cancer therapy, study completion, or early study termination by the Sponsor.

5.4.3.2.1. EORTC QLQ-C30

The EORTC QLQ-C30 was developed to be an "integrated, modular approach for evaluating the quality of life of patients participating in international [oncology] clinical trials." It is comprised of 5 functional scales (physical, role, cognitive, emotional, and social), three symptom scales (fatigue, pain, and nausea and vomiting), a global health status scale, a number of single items assessing additional symptoms commonly reported by cancer patients (dyspnea, loss of appetite, insomnia, constipation, and diarrhea), and perceived financial impact of the disease. Validation experience documents that the scales assess distinct components of the health-related quality of life construct.

Functional scales:

- Physical functioning: (1 - ((Q1+Q2+Q3+Q4+Q5)/5 -1)/3) * 100
- Role functioning: (1 - ((Q6+Q7)/2-1)/3) * 100
- Emotional functioning: (1 - ((Q21+Q22+Q23+Q24)/4-1)/3) * 100
- Cognitive functioning: (1 - ((Q20+Q25)/2-1)/3) * 100
- Social functioning: (1 - ((Q26+Q27)/2-1)/3) * 100

Global health status:

- Global health status/QoL: ((Q29+Q30)/2-1)/6 * 100

Symptom scales/items:

- Fatigue: ((Q10+Q12+Q18)/3-1)/3 * 100
- Nausea and vomiting: ((Q14+Q15)/2-1)/3 * 100
- Pain: ((Q9+Q19)/2-1)/3 * 100
- Dyspnea: ((Q8-1)/3 * 100
- Insomnia: (Q11-1)/3 * 100
- Appetite loss: (Q13-1)/3 * 100
- Constipation: (Q16-1)/3 * 100
- Diarrhea: (Q17-1)/3 * 100
- Financial difficulties: (Q28-1)/3 * 100

Baseline and change from baseline in EORTC QLQ-C30 global health status/quality of life (QoL) composite scale data and the remaining EORTC QLQ-C30 scale data will be summarized by time point using descriptive statistics. In addition, the percentage of patients demonstrating a

clinically meaningful deterioration (defined as a 10-point change from baseline) will be presented for each scale at each assessment time point. Percentages will be based on number patients assessed at each assessment time point.

A linear mixed effect regression model with an appropriate covariate structure will be used to fit the longitudinal data. In the linear mixed regression model, stratification factors used in the randomization will be included as fixed factors. The model will also include time, study arm, and study arm by time interaction. The following covariance structure will be explored: Unstructured (TYPE=UN), compound symmetry structure (TYPE=CS) and first-order autoregressive (TYPE=AR(1)). The covariance structure resulting in model convergence and the lowest Bayesian Information Criterion (BIC) will be used for analysis. Change from baseline scores from all post-baseline visits will be compared between the two study arms based on the linear mixed effects model. Subjects without baseline score or any post-baseline score will be excluded from the linear mixed effects regression analysis.`

5.4.3.2.2. EORTC QLQ-BIL21

The EORTC QLQ-BIL21 is a validated tool for the assessment of QoL in patients with CCA and cancer of the gallbladder. It has five multi-item scales: eating (four items); jaundice (three items); tiredness (three items); pain (four items) and anxiety (four items) and three single items about treatment side effects, drainage tubes/bags and worried about losing weight.
Multi-item scales: [Need EORTC BIL21 Scoring Manual; scoring algorithm below are based on EORTC QLQ-C30 Scoring methods with 0-to-100 linear transformation]

- Eating: (1 - ((Q1+Q2+Q3+Q4)/4 -1)/3) * 100
- Jaundice: (1 - ((Q5+Q6+Q7)/2-1)/3) * 100
- Tiredness: (1 - ((Q8+Q9+Q10)/3-1)/3) * 100
- Pain: (1 - ((Q11+Q12+Q13+Q14)/4-1)/3) * 100
- Anxiety: (1 - ((Q15+Q16+Q17+Q18)/4-1)/3) * 100

Baseline and change from baseline in EORTC QLQ-BIL21 will be summarized by time point using descriptive statistics for each cohort. In addition, the percentage of patients demonstrating a clinically meaningful deterioration (defined as a 10-point change from baseline) will be presented for each scale at each assessment time point. Percentages will be based on number patients assessed at each assessment time point. The changes values from baseline to post-baseline and shift evaluation from baseline to post-baseline will be summarized.

The changes values from baseline to post-baseline and shift evaluation from baseline to post-baseline will be summarized. The similar mixed linear mixed effect regression model descried in 5.4.3.2.1 will be used.

5.4.4. Subgroup Analyses

To evaluate of the impact of demographics and baseline characteristics on efficacy, subgroup analyses on efficacy endpoints including, but not limited to, PFS, OS, ORR and DOR will be performed for subgroups including, but not limited to, those defined below:

- Geographic Region (Asia vs. rest of the world).
- Age categorization (< 65, ≥65).
- Gender (male, female).
- Race (White, African American, Asian, Other).
- Baseline ECOG PS (0 vs. 1).

For PFS/OS, estimate of hazard ratio along with its two-sided 95% CI will be presented for subgroups defined above.

For ORR, the proportion of patients with response along with the treatment difference will be summarized for subgroups defined above along with its 95% exact CI.

The consistency of the treatment difference among these sub-populations will be examined by forest plots.

5.4.5. Handling of Missing Data

For time to event endpoints such as PFS and OS, detail event definition and censoring rule are described in Table 2 and Table 3, respectively.

For response endpoints such as ORR and DCR, patients without any post-baseline assessment will be considered non-responders.

For DOR, patients with no documented progression after CR/PR will be censored at the last date of adequate assessment. Patients who have never responded will not be included in this analysis.

For PRO instruments, missing values will be imputed for missing items by "assuming that the missing items have values equal to the average of those items which are present" for any scale in which at least half the items are completed. A scale in which less than half of the items are completed will be treated as missing. A questionnaire will be considered as received if at least one of the 15 scales is non-missing (after imputation). All scales and single items are scored on a categorical scale and linearly transformed to 0-to-100 scales with higher scores for a functional scale representing higher levels of functioning, higher scores for the global health status/quality of life representing higher levels of global health status/quality of life, and higher scores for a symptom scale representing higher level of symptoms.

5.4.6. Multiplicity Adjustment

To adjust for multiple testing of the secondary endpoints, a fixed sequence testing procedure will be used. The following endpoints will be tested in the order listed below:

- ORR
- DCR
- OS
- PFS per Investigator assessment

If the study meets its primary efficacy endpoint of prolonged PFS in the treatment arm, a formal statistical test of ORR between the two arms will be performed. After achieving statistical significance at $\alpha = 0.05$ (2-sided) with primary endpoint PFS, the secondary endpoint ORR will be compared. If the study does not meet its primary endpoint, this test will not be performed.

If the ORR comparison achieves statistical significance, DCR will be tested at a two-sided significant level of 0.05. OS and PFS per investigator will also be tested sequentially if the previous comparison reaches its statistical significance.

The p-values generated from other analyses outside this hierarchical testing procedure are not inferential, for descriptive purpose only.

5.5. Pharmacokinetic Analysis

The concentration vs. time data obtained in this study will be combined with data from other studies in the clinical development program to develop a Pop PK model. This model will be used to evaluate the effects of intrinsic and extrinsic covariates on the PK of futibatinib and to estimate measures of individual exposure (such as steady-state peak, trough and area under the plasma drug concentration-time curve). Model estimated exposures will be used for exposure-response analyses of selected efficacy and safety endpoints.

The analysis procedure of the PopPK and exposure-response analyses will be detailed in a separate analysis plan, and the results of the analyses will be reported separately.

The list of futibatinib plasma concentrations will be provided.

5.6. Pharmacodynamic/Biomarker Analysis

There is no PD biomarkers for this study except, probably, circulating CA19.9 data which will be collected from local results. If enough data is collected, % change of on-treatment levels from baseline will be calculated as a patient monitoring biomarker.

Tissue *FGFR2* rearrangement status reported to sites will be collected from the EDC but will be also obtained from the central lab. Correlation of clinical outcome with different types of *FGFR2* rearrangements/fusions will be assessed.

Other tissue NGS results will be collected from the central lab for exploratory statistical analysis which can be handled outside this formal SAP.

Data from circulating free DNA (cfDNA), if analyzed, will be collected from the central lab for exploratory statistical analysis which can be handled outside this formal SAP.

5.7. Safety Analyses

All safety analyses will be performed on the all treated population. No inferential statistics are planned.

The safety analysis will be performed by treatment arms and overall on the as-treated patients (all treated population).

AEs will be coded according to the MedDRA (Version 23.1 or later) terminology and the severity of the toxicities will be graded according to NCI CTCAE Version 5.0, where applicable.

5.7.1. Extent of Exposure

5.7.1.1. Administration of Study Drug

The following parameters will be summarized by treatment group:

- Duration of treatment
- Number of doses received
- Cumulative dose
- Relative dose intensity

Duration of treatment will be calculated as: (Date of last dose – Date of first dose + 1).

Cumulative dose is sum of the doses administered to a subject during the treatment period.

Relative dose intensity will be calculated as cumulative dose (dose received) divided by dose planned (taking the dosing cycle into consideration)

A by-patient listing of dosing of study medication and a listing of batch number will be also provided.

5.7.1.2. Modification of Study Drug

The number and percentage of patients with at least one dose interruption along with reason for the interruptions will be summarized by treatment group.

The number and percentage of patients with at least one dose reduction and reason for the dose reduction will be summarized by treatment group.

By-patient listings of dose interruptions and dose reductions will be also provided.

5.7.2. Adverse Events

5.7.2.1. Deaths

Deaths and reasons for death will be summarized by treatment group for all deaths on treatment or within 30 days of last dose received.

A by-patient listing of all deaths will be provided.

5.7.2.2. Adverse Events

An AE is any untoward medical condition that occurs in patients while participating in this clinical trial. A treatment-emergent AE (TEAE) is defined as an AE that is starting or worsening at the time of or after the first dose of study drug administration and within 30 days after the last dose of study drug, and does not necessarily have a causal relationship to the use of the study drug.

Treatment-related AEs (TRAEs) will be reported in the table.

The following summary tables will be generated:

1. Summary of AEs with the number and percentage of patients reporting AEs, serious AEs, grade ≥ 3 AEs, AEs leading to study drug discontinuation, interruption and dose modification, and AEs with the outcome of death.
2. Summary of TEAEs with the number and percentage of patients reporting TEAEs, serious TEAEs, grade ≥ 3 TEAEs, TEAEs leading to study drug discontinuation, interruption and dose modification, and TEAEs with outcome of deaths.
3. Summary of TRAEs with the number and percentage of patients reporting TRAEs, serious TRAEs, grade ≥ 3 TRAEs, TRAEs leading to study drug discontinuation, interruption and dose modification, and TRAEs with outcome of deaths.
4. Summary of AEs by worst CTC Grade (grade 1, 2, 3, 4, 5, any grade, Grade ≥ 3) presented by SOC and Preferred Term (PT).
5. Summary of TEAEs by worst CTC grade (Grade 1, 2, 3, 4, 5, any grade, Grade ≥ 3) presented by SOC and PT.
6. Summary of TRAEs by worst CTC grade (Grade 1, 2, 3, 4, 5, any grade, Grade ≥ 3) presented by SOC and PT.
7. Grade ≥ 3 AEs by worst CTC Grade (grade 3, 4, 5, 3-5) presented by SOC and PT.
8. Grade ≥ 3 TEAEs by worst CTC Grade (grade 3, 4, 5, 3-5) presented by SOC and PT.
9. Grade ≥ 3 TRAEs by worst CTC Grade (grade 3, 4, 5, 3-5) presented by SOC and PT.
10. Summary of AEs leading to death by SOC and PT.
11. Summary of TEAEs leading to death by SOC and PT.
12. Summary of TRAEs leading to death by SOC and PT.
13. Summary of the most common AEs (≥10% incidence) and median time to first AE by SOC and PT.
14. Summary of the most common TEAEs (≥10% incidence) and median time to first TEAE by SOC and PT.
15. Summary of the most common TRAEs (≥10% incidence) and median time to first TRAE by SOC and PT.

5.7.2.3. Serious Adverse Events

Serious adverse events (SAEs) will be summarized:

- Summary of SAEs by worst CTC Grade (grade 1, 2, 3, 4, 5, any grade, Grade ≥ 3) presented by SOC and PT
- Summary of treatment-related SAEs by worst CTC Grade (grade 1, 2, 3, 4, 5, any grade, Grade ≥ 3) presented by SOC and PT

A by-patient listing of SAEs will also be provided.

5.7.2.4. Adverse Events Leading to Discontinuation of Study Drug

AEs leading to discontinuation will be summarized by treatment group:

- Overall summary of AEs leading to discontinuation by worst CTC Grade (grade 1, 2, 3, 4, 5, any grade, Grade 3-5) presented by SOC/PT
- Overall summary of drug-related AEs leading to discontinuation by worst CTC grade (Grade 1, 2, 3, 4, 5, any grade, Grade 3-5) presented by SOC/PT

A by-patient listing of AEs leading to discontinuation will also be provided.

5.7.2.5. Adverse Events Leading to Dose Modification

AEs leading to dose modification (including dose interruption or reduction) will be summarized by treatment group:

- Overall summary of AEs leading to modification by worst CTC Grade (grade 1, 2, 3, 4, 5, any grade, Grade 3-5) presented by SOC/PT
- Overall summary of drug-related AEs leading to discontinuation by worst CTC grade (Grade 1, 2, 3, 4, 5, any grade, Grade 3-5) presented by SOC/PT

A by-patient listing of AEs leading to modification will also be provided.

5.7.2.6. Multiple Events

The following summary tables will be provided:

- A table showing the total number and rate (exposure adjusted) of occurrences for all AEs.

In addition, the rate (exposure adjusted) and its 95% CI evaluated for different time intervals will be displayed graphically for each treatment group. This analysis will be limited to the rate of all AEs and all drug-related AEs. No formal comparisons will be made between treatment groups.

Listing displaying the unique instances of all AEs (that is, after duplicates have been eliminated and overlapping and contiguous occurrences of the same event (same PT) have been collapsed) will be provided.

5.7.3. Clinical Laboratory Evaluations

5.7.3.1. Hematology

The following will be summarized by treatment group as worst CTC grade on-treatment per patient and as shift table of worst on-treatment CTC grade compared to baseline CTC grade per patient: hemoglobin (HB), platelets, white blood counts (WBC), absolute neutrophils count (ANC) and lymphocyte count (LYMPH).

A by- patient listing of these laboratory parameters will be provided.

5.7.3.2. Chemistry

The following will be summarized by treatment group as worst CTC grade on-treatment per patient and as shift table of worst on-treatment CTC grade compared to baseline CTC grade per patient: ALT, AST, alkaline phosphatase (ALP), inorganic phosphorus glucose, total bilirubin, creatinine. sodium (high and low), potassium (high and low), calcium (high and low) and magnesium (high and low).

A by- patient listing of these laboratory parameters will be provided.

5.7.3.3. Additional Analyses

The number of patients with the following laboratory abnormalities from on-treatment evaluations will be summarized by treatment group:

- ALT or AST > 3 x ULN, > 5 x ULN, > 10 x ULN and > 20 x ULN
- Total bilirubin > 2 x ULN
- Concurrent (within 1 day) ALT or AST > 3 x ULN and total bilirubin > 2 x ULN
- Concurrent (within 30 days) ALT or AST > 3 x ULN and total bilirubin > 2 x ULN
- ALT or AST >3 x ULN and total bilirubin >2 x ULN bilirubin and alkaline phosphatase <2 x ULN or missing
- ALT or AST >3xULN total bilirubin >2 x ULN bilirubin and alkaline phosphatase <2 x ULN

Scatterplots will be produced for the following hepatic laboratory parameters:

- Total bilirubin peak vs. AST peak
- Total bilirubin peak vs. ALT peak

On-treatment peak total bilirubin and on-treatment peak AST/ALT may or may not happen on the same day of liver testing. A by-patient listing of these specific abnormalities will be provided.

5.7.4. Vital Signs

Descriptive statistics for vital signs parameters (diastolic and systolic blood pressure, pulse, respiration rate, temperature, <weight (if collected> and changes from baseline will be presented by visit and treatment group (and/or dose).

5.7.5. Electrocardiograms

Descriptive statistics for ECG parameters and changes from baseline will be presented by visit and treatment group (and/or dose).

Shift tables will present changes from baseline in ECG interpretation (categorized as normal; abnormal, not clinically significant; and abnormal, clinically significant) to end of treatment (or end of phase or by visit).

In addition, the number (percentage) of patients with at least 1 post-baseline abnormal ECG result in QTc Fridericia (and/or QTc Bazett) during the on-treatment period will be summarized. Clinically abnormal ECG results in QTc Fridericia (and/or QTc Bazett) will be categorized as follows:

Absolute QTc interval prolongation:

- QTc interval > 450 ms
- QTc interval > 480 ms
- QTc interval > 500 ms

Change from baseline in QTc interval:

- QTc interval increases from baseline > 30 ms
- QTc interval increases from baseline > 60 ms

5.8. Interim Analyses

No interim analysis of efficacy is planned. An independent Data Monitoring Committee (DMC) will be established prior to the first interim analysis to monitor data on an ongoing basis to ensure the continuing safety of the patients enrolled in this study. Interim analyses for safety monitoring will occur when approximately 25%, 50%, and 75% of the planned total number of patients have been treated. DMC will meet to assess the cumulative safety data and recommend study continuation, discontinuation, or study modification. A description of the roles and responsibilities of the DMC and details of the review processes are provided in a separate DMC charter.

Prior to final study unblinding, the Sponsor and external unblinded statistician will not be involved in any discussions regarding modifications to the protocol, statistical methods,

6. CHANGES IN PLANNED ANALYSIS

There have been no changes to the initially planned analyses as of the finalization of this SAP.

7. DEFINITIONS AND CONVENTIONS FOR DATA HANDLING

This section should contain detailed definitions and conventions, including data handling rules that will be used to perform statistical analyses.

7.1. Baseline Period

Baseline evaluations or events will be defined as evaluations or events that occur before the date and time of the first dose of study drug. Evaluations on the same date and time of the first dose of study drug will be considered as baseline evaluations.

In cases where the time (onset time of event or evaluation time and dosing time) is missing or not collected, the following definitions will apply:

- Pre-treatment AEs will be defined as AEs with an onset date prior to but not including the day of the first dose of study drug
- Baseline evaluations (laboratory tests, pulse oximetry and vital signs) will be defined as evaluations with a date on or prior to the day of first dose of study drug

If there are multiple valid assessments, the assessment that is closest to day (and time if collected) of the first dose of study drug will be used as the baseline in the analyses. If multiple assessments are collected on the same date (and time, if collected), the assessment with the latest database entry date (and time, if collected) will be considered as baseline.

7.2. Post-Baseline Period

On-treatment AEs will be defined as AEs with an onset date and time on or after the date and time of the first dose of study drug (or with an onset date on or after the day of first dose of study drug if time is not collected or is missing). For patients who are off study drug, AEs will be included if event occurred within a safety window of 30 days after the last dose of study drug. No "subtracting rule" will be applied when an AE occurs both pre-treatment and post-treatment with the same preferred term and grade.

On-treatment evaluations (laboratory tests, pulse oximetry and vital signs) will be defined as evaluations taken after the day (and time, if collected and not missing) of first dose of study drug. For patients who are off study drug, evaluations should be within a safety window of 30 days after the last dose of study drug.

7.3. Other Missing Data Handling Rules

Missing data will not be imputed in listings. The listings will only present the data recorded on the original Case Report Form.

If an AE has a completely missing onset date, then the AE will be considered a TEAE. A medication with a completely missing start date is considered a prior medication. A medication with a completely missing stop date is considered a concomitant medication.

If an AE or a medication has a partial missing start date or stop date, the following rules in Table 5 will be used to determine whether it is a TEAE, or a prior or concomitant medication. The rules are also used for determining the cycle of AE and concomitant medication. The derived date is used for determining TEAEs, the cycle of AE, and concomitant medication.

Table 5: Partial Date Imputation Rule for AE or Medication

Partial Missing Start or Stop Date	Derived Start Date	Imputed Stop Date
Missing month and day, and the year is present	January 1 of that year or first dose date if the year is the same as the year of first dose date	December 31 of that year
Missing day, but year and month are present	First day of that month or first dose date if the year and month are the same as the year and month of first dose date	Last day of that month
Missing month, but year and day are present	Missing month derived as January or same as first dose month if the year is same as the year of first dose.	Missing month imputed as December

For the medical history, prior surgery, prior radiotherapy, and prior systemic cancer therapies:

- A partial missing start dates will be imputed as below
  - Missing year- no imputation, data left missing
  - Missing day and month - January 1 will be assigned to the missing fields.
  - Missing month only - Treat day as missing and replace both month and day according to the above procedure.
  - Missing day only - Assign first of the month to the missing day

- A partial missing stop dates will be imputed as below
  - Missing year: no imputation, date left missing
  - Month is missing and year is prior to year of first dose of study drug: impute 'December'.
  - Month is missing and year is the same as the year of the first dose of study drug: impute same month as in first dose date of study drug.
  - Day is missing: impute 'last day of that month'. If results in a date ≥ the date of the first dose of study drug impute day as the day prior to date of first dose of study drug

8. REFERENCES

- Clopper CJ, Person ES. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometricka 1934; 26, 404-13.
- Brookmeyer R, Crowley J. A confidence interval for the median survival time. Biometrics. 1982;38:29-41.
- Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. John Wiley & Sons, Inc., New York, 1980. Xi + 321 pp.
- Westfall PH, Krishen A. Optimal weighted, fixed sequence and gatekeeper multiple testing procedure, J Stat Plan Inference 2001; 99:25-40.
- FDA Guidance on Conduct of Clinical Trials of Medical Products during COVID-19 Pandemic, Updated on March 27, 2020