BASILEA PHARMACEUTICA

STATISTICAL ANALYSIS PLAN

A PIVOTAL STUDY OF DERAZANTINIB IN PATIENTS WITH INOPERABLE OR ADVANCED INTRAHEPATIC CHOLANGIOCARCINOMA AND *FGFR2* GENE FUSIONS OR *FGFR2* GENE MUTATIONS OR AMPLIFICATIONS

DZB-CS-301 (formerly ARQ 087-301)

LABCORP STUDY ID: 000000154565

DOCUMENT VERSION: Final

DOCUMENT DATE: 23 August 2022

AUTHOR: [redacted]

LABCORP CLINICAL DEVELOPMENT LIMITED

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

AE	Adverse event
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical (drug classification system)
BICR	Blinded Independent Central Review
BMI	Body mass index
CCG	Clinical change group
cm	centimeter
CR	Complete response
CT	Computed tomography
CTCAE	Common Terminology Criteria for Adverse Events
ctDNA	Circulating tumor deoxyribonucleic acid
DCR	Disease control rate
DMC	Data monitoring committee
DoR	Duration of response
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EOT	End of treatment
EORTC	European Organization for Research and Treatment of Cancer
EQ-5D (5L)	EuroQol five-dimension scale five-level
FISH	Fluorescence in situ hybridization
GA	Genetic aberration
G-SET	Global Self Evaluated Transition
H_0	Null hypothesis
H_A	Alternative hypothesis
HRQoL	Health-related quality of life
HTI	Health Transition Index
iCCA	Intrahepatic cholangiocarcinoma
IEC	Independent Ethics Committee
IPD	Important protocol deviation
IRB	Institutional Review Board
ITT	Intent-to-treat
kg	Kilogram
kg/m^2	Kilogram per meter squared
LDH	Lactate dehydrogenase
LLN	Lower limit of normal
LFT	Liver function test(s)
m^2	Meter squared
max	Maximum
MedDRA	Medical Dictionary for Regulatory Activities
min	Minimum

MID	Minimally important difference
mITT	Modified intent-to-treat
mg	Milligram
mmHg	Millimeter of mercury
MRI	Magnetic resonance imaging
n	Number of available observations
NE	Not evaluable
NCI CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
NGS	Next-generation sequencing
OR	Objective response
ORR	Objective response rate
OS	Overall survival
PD	Progressive disease
PFS	Progression free survival
PFS3	Progression free survival at 3 months
PK	Pharmacokinetics
PopPK	Population pharmacokinetics
PP	Per protocol
PR	Partial response
PT	Preferred term
Q1	First quartile
Q3	Third quartile
QD	Once daily
QTc	Corrected QT Interval
RECIST	Response Evaluation Criteria in Solid Tumors
RS	Raw score
SAE	Serious adverse event
SAP	Statistical analysis plan
s.d.	Standard deviation
SD	Stable disease
SOC	System organ class
TEAE	Treatment-emergent adverse event
TTP	Time to progression
ULN	Upper limit of normal
VAS	Visual analog scale
WHODRUG	Who Drug Dictionary

1 OVERVIEW AND INVESTIGATIONAL PLAN

This statistical analysis plan (SAP) was created based on Protocol DZB-CS-301, Version 9.0, Amendment 8.0 dated 17 November 2020. It provides a comprehensive and detailed description of the planned statistical analysis for the above mentioned protocol.

1.1 STUDY DESIGN

This is a multi-center, open label, single arm, Phase 2 study evaluating derazantinib in adult patients with inoperable or advanced intrahepatic cholangiocarcinoma (iCCA) and *FGFR2* genetic aberrations (GAs).

- **Substudy 1** enrolls patients with *FGFR2* fusions.
- **Substudy 2** enrolls patients with *FGFR2* mutations or amplifications

Substudy 1 was considered pivotal and will enroll approximately 100 patients to determine the objective response rate (ORR).

Substudy 2 will use a Simon's two-stage design, with approximately 15 modified intent-to-treat (mITT) - evaluable patients in Stage 1, and an additional 28 mITT-evaluable patients if the study proceeds to Stage 2.

Patient eligibility will be assessed in two steps: 1) Molecular pre-screening to confirm *FGFR2* GA status (*FGFR2* fusions, or *FGFR2* mutations/amplifications; see Table 1 and Table 2, and Appendix 8 from the protocol) using validated/approved genetic testing devices; and 2) clinical screening procedures to confirm study treatment eligibility is confirmed by the tumor's *FGFR2* GA status.

Molecular pre-screening for *FGFR2* fusions in Substudy 1 may be based on local or central testing. If pre-screening is performed based on a local test, then a central confirmation by the *FGFR2* break-apart fluorescence in situ hybridization (FISH) Probe Kit test is required for Substudy 1 (*FGFR2* fusions). Molecular pre-screening for *FGFR2* mutations/amplifications in Substudy 2 should be based on next-generation sequencing (NGS) testing performed or commissioned by the respective study site.

Patients may be prescreened for the tumor's *FGFR2* GA status (prior to initiation of or during ongoing systemic therapy). If the patient is still receiving prior systemic therapy, clinical screening procedures will be delayed until radiographically confirmed disease progression or intolerance to the ongoing systemic therapy is documented. After the study treatment eligibility is confirmed, patients will be enrolled and treated with continuous 300 mg once daily (QD) of derazantinib capsules. A treatment cycle is defined as 28 days. Patients will receive treatment with derazantinib capsules until death, radiographic disease progression, unacceptable toxicity, or until another of the specified criteria is met for stopping therapy.

If the locally-documented *FGFR2* gene fusion positive status was tested and not confirmed by FISH (Substudy 1) by the central laboratory designated by the Sponsor after the commencement of study treatment, these patients will be assessed on a case-by-case basis.

It is expected that most patients will receive between 1 and 8 months of treatment with derazantinib capsules. Dose delays and/or reductions will be allowed when derazantinib-related toxicity is observed.

During the treatment period, patients will be evaluated every 2 weeks for the first cycle (Cycle 1 Days 1 and 15), and once every cycle thereafter (Day 1 of each cycle).

Tumor measurements will be done at Screening (within 28 days prior to the first dose of derazantinib), once every 8 weeks (two cycles) from the day of the first dose for the first 6 cycles and once every 12 weeks (three cycles) thereafter. For patients with partial response (PR) or complete response (CR), the Investigator should perform a confirmation tumor measurement 4 to 5 weeks after the scan showing PR or CR. For patients with progressive disease (PD) per Investigator assessment, a central radiology reviewer confirmation should be received prior to the patient's discontinuation from the study drug if progression is seen on the first or second on-treatment scan. Patients who discontinue study drug for a reason other than radiographic disease progression, withdrawal of consent, death, or loss to follow-up, and have not had their end of treatment (EOT) visit, should continue tumor evaluation visits if possible every 8–12 weeks until they start another anti-cancer therapy, experience disease progression, withdraw consent, die, or are lost to follow-up.

Pharmacodynamic assessment will include evaluation of tumor markers (CA19.9, CA125, CEA), biomarkers (FGF19, FGF21, FGF23), and cell-free circulating tumor deoxyribonucleic acid (ctDNA). Blood samples for tumor markers will be collected for all enrolled patients. Blood samples for PD assessments will be collected only from patients enrolled after completion of the interim analysis, and subject to the granting of appropriate regulatory and Institutional Review Board (IRB)/Independent Ethics Committee (IEC) approval. These blood samples will be collected on Day 1 of the first cycle, every 8 weeks (two cycles) for the first six cycles, once every 12 weeks (three cycles) thereafter, and at the End of Treatment visit. All collected samples will be evaluated in batches either during or at the end of the study.

Archival tumor tissue samples will be obtained for all patients to enable additional molecular analyses with regard to the identified GAs and markers predictive of response at the laboratories selected by the Sponsor. In addition, further information on the molecular status of the tumor, including available full NGS reports will be collected.

To determine the population pharmacokinetics (PopPK) parameters of derazantinib, blood samples will be collected on Day 1 and Day 15 of Cycle 1, and on Day 1 of Cycles 2, 3, and 4. Exploratory assessments of metabolites of derazantinib will also be investigated from these PK plasma samples.

All patients will be scheduled for sparse PK sampling. A subset of up to 20 patients (PK subgroup) will be asked to participate in rich PK sampling and/or urinary PK sampling.

Health-related quality of life (HRQoL) and symptom response will be measured using the QLQ-C30, QLQ-BIL21, and the EuroQol five-dimension scale five-level (EQ-5D) (5L). The Global Self Evaluated Transition (G-SET) / Health Transition Index (HTI) is a single item, and will be used as an external anchor to determine the minimal important difference of the EQ-5D visual analog scale (VAS), European Organization for Research and Treatment of Cancer (EORTC) QLQ-C30, and QLQ-BIL21 scales.

Safety follow-up will be conducted at least 30 days after the administration of the last dose of study drug. Safety follow-up will include collection of adverse events (AEs) and changes in concomitant medication. Survival follow-up will start the day of the last dose of derazantinib; it will continue until the study has completed or other discontinuation criteria are met.

Assessments and procedures during all cycles are performed according to the Schedule of Assessments in Table 1 unless otherwise indicated.

Note: C1D1 denotes Cycle 1, Day 1, CXDX denotes Cycle X, Day X.

Table 1 Schedule of Assessments

Tests & Procedures / Visit Name	Pre-Screening Visit	Screening Visit	Cycle 1		Cycle 2+	End of Treatment Visit	30-day Safety Follow-up	Overall Survival Follow-up
Day	NA	0	1	15	1	Within 7 days after last dose of derazantinib or decision to permanently stop dosing	30+ days after the last dose of derazantinib	At least every 3 months from date of last dose of derazantinib
Window	NA	-28 to -1	0	± 3 days	± 3 days	NA	NA	±14 days
Written informed consent	X[1]	X[1]						
Medical history	X	X						
QLQ-C30, QLQ-BIL21, EQ-5D, G-SET/HTI questionnaires (prior to seeing physician)			X[2]		X[2]	X[2]	X[2]	
Physical examination		X	X	X	X	X	X	
Complete opthalmological examination		X			X[3]	X	X[3]	
ECOG performance status	X	X	X	X	X	X	X	
Vital signs[4]		X	X	X	X	X	X	
12-Lead ECG, in triplicate		X[5]	X[5]	X[5]	X[5]	X[5]		
Blood for clinical blood tests[6]		X	X	X	X	X	X	
Blood for serum pregnancy test		X[7]	X[7]		X[7]	X[7]	X[7]	
Blood for tumor markers, biomarkers, and ctDNA			X[8]		X[8]	X[8]		
Blood for pharmacokinetics			X[9]	X[9]	X[9]			
Urine for Pharmacokinetics					X[10]			
Tumor's genomic status[11]	X							
Tumor measurement and staging		X[12]			X[13]	X[14]		
Concomitant medications		X[15]	X[15]	X	X	X	X	
Adverse events assessment			X	X	X	X	X	
Derazantinib capsule dispensing[16]			X		X			
Derazantinib capsule accountability					X	X		
Survival contact[17]								X

[1] Consent can be obtained greater than 28 days prior to first dose and does not have to be repeated unless an updated, approved consent form is available. A patient can consent for tissue analysis and the full study separately or at the same time.

[2] EQ-5D, EORTC QLQ-C30 and QLQ-BIL21 questionnaires should be administered on Day 1 of the first cycle, every 8 weeks (two cycles) for the first six cycles (C3D1, C5D1, C7D1), once every 12 weeks (three cycles) thereafter (C10D1, C13D1, etc.), at the End of Treatment visit and at the 30 day Safety Follow-up. G-SET/HTI questionnaire should be administered on Day 1 of Cycle 3 and Day 1 of Cycle 5. EQ-5D should be completed first, followed by EORTC QLQ-C30, QLQ-BIL21 and G-SET/HTI, when applicable.

[3] Complete opthalmological examination will be performed at C2D1, C3D1, C4D1, C5D1, at the End of Treatment visit, and at the 30-day Safety Follow-up visit, and if clinically indicated (see Section 6.3 from the protocol).

[4] Vital signs include weight, temperature, blood pressure, respiration rate, and pulse. At the Screening visit, height will also be measured (see Section 6.2 from the protocol).

[5] 12-lead ECG in triplicate is required at Screening., on Day 1 and Day 15 of Cycle 1 (pre-dose, and approximately 6–8 hours after the dosing), on Day 1 of Cycle 2 (pre-dose), on Day 1 of Cycle 3 (pre-dose, and approximately 6–8 hours after the dosing), on Day 1 of Cycle 4 and all subsequent cycles (pre-dose), and at the End of Treatment visit (see Section 6.4 from the protocol).

[6] Clinical safety blood samples will be forwarded to a central laboratory designated by the Sponsor for testing (see Section 6.5 from the protocol).

[7] A serum pregnancy test, if applicable, is required at Screening within 72 hours prior to dosing, and at the End of Treatment visit. In addition, serum- or urine pregnancy testing will be performed monthly (on Day 1 of each treatment cycle) while the patient receives study treatment, at the 30-day Safety Follow-up visit, and until 120 days after the last administration of study drug. (see Section 8.8.2 from the protocol).

[8] Blood samples for tumor markers, biomarkers, and ctDNA are only collected on Day 1 of Cycle 1 at pre-dose and every 8 weeks (two cycles) for the first six cycles (C3D1, C5D1, C7D1), once every 12 weeks (three cycles) thereafter (C10D1, C13D1, etc.), and at the End of Treatment visit. Blood samples for biomarkers and ctDNA will be collected only from patients enrolled after completion of the interim analysis, and subject to the granting of appropriate regulatory and IRB/IEC approval (see Section 6.7 and Appendix 3 from the protocol).

[9] Blood for sparse PK sampling is collected on C1D1, C1D15 and C3D1 at pre-dose (− 1 hour) and 6–8 hours post-dose. Blood for PK is collected on Day 1 of Cycles 2 and 4 at pre-dose (− 1 hour) (see Section 6.6 and Appendix 3). Blood for rich PK sampling for the PK/biomarker subgroup is collected on C1D1 and C2D1 pre-dose (− 1 hour), and 1 (± 5 minutes), 2 (± 5 minutes), 4 (± 15 minutes), 6 (± 15 minutes), 8 (± 30 minutes), 10 (± 30 minutes), 12 (± 30 minutes), and 24 hours after the daily dose of derazantinib (within 1 hour prior to the next dose), on C1D15 and C3D1, pre-dose (− 1 hour) and between 6 and 8 hours after the daily dose of derazantinib, and on C4D1 pre-dose (− 1 hour); the 10 and 12 h time-points are optional (see Appendix 3 from the protocol).

[10] In the rich PK sampling subset for urinary excretion: Urine for PK is collected on Day 1 of Cycle 2, starting from daily dose administration and until 24 h after dosing (immediately prior to the next dose) (see Appendix 3 from the protocol).

[11] The patient must test positive for *FGFR2* fusion (Substudy 1) or *FGFR2* mutation or amplification (Substudy 2). Details of FISH and NGS testing for this purpose are provided in Section 6.8, and summarised in Table 1 and Table 2 from the protocol.

[12] Tumor imaging assessments must be within 28 days prior to the first dose (see Section 6.9 from the protocol).

[13] Tumor measurement (CT/MRI scan of the chest, abdomen, and pelvis) and staging will be done every 8 weeks (two cycles) for the first six cycles (C3D1, C5D1, C7D1) and once every 12 weeks (three cycles) thereafter (C10D1, C13D1, etc.). Post-dose BS / WBMRI will be performed only if clinically indicated. If a scan shows CR/PR, a confirmation scan must be performed 4-5 weeks after the last scan was performed (see Section 6.9 from the protocol).

[14] Tumor measurement and staging will be performed at the End of Treatment visit only if the previous scan was not done within four weeks (28 days) prior to the End of Treatment visit or if the previous scan did not show radiographic disease progression (see Section 6.9 from the protocol).

[15] All medications taken within 30 days prior to the first dose are to be recorded.

[16] To avoid unnecessary waste of derazantinib capsules, in cases where treatment was interrupted and/or dose was reduced, the patient can continue dosing from the previously dispensed bottle until the next drug dispensing visit where re-supply is needed to maintain the protocol dosing regimen.

[17] Survival contact can be in person, via phone, or, where applicable, by checking regional/national death registries. All patients and/or family will be contacted at 3 month intervals (± 14 days) to record the patient status as Alive (date); Dead (date); Alive, but withdrew consent for further follow up; Lost to Follow Up. Survival updates may be made more often than every 3 months if the patient is seen at the investigational site for other reasons and for study level survival sweep(s) (see Section 5.6 from the protocol).

1.2 OBJECTIVES

1.2.1 Primary objective

Substudy 1

- To evaluate the anti-cancer activity of ORR by central radiology review as per Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1 in patients with inoperable or advanced iCCA whose tumors harbor *FGFR2* fusions and who received at least one prior regimen of systemic therapy.

Substudy 2

- To evaluate the anti-tumor activity of derazantinib by progression-free survival (PFS) at 3 months (PFS3) based on survival status or central radiology review (RECIST version 1.1) in patients with inoperable or advanced iCCA whose tumors harbor *FGFR2* mutations or amplifications, and who received at least one prior regimen of systemic therapy.

1.2.2 Secondary objectives

Substudy 1 and Substudy 2

- To evaluate PFS by central radiology review and overall survival (OS)
- To evaluate duration of response (DoR) by central radiology review
- To evaluate the safety profile (toxicities) of derazantinib in this patient population
- To evaluate changes, and assess the minimally important difference, in HRQoL and symptom response from baseline using the EORTC QLQ-C30, QLQ-BIL21, G-SET / HTI, and the EQ-5D VAS

Substudy 2

- To evaluate the anti-cancer activity by ORR by central radiology review as per RECIST version 1.1 in patients with inoperable or advanced iCCA whose tumors harbor *FGFR2* mutations or amplifications, and who received at least one prior regimen of systemic therapy

1.2.3 Exploratory objectives

The exploratory efficacy objectives of this study are:

- To evaluate disease control rate (DCR) by central radiology review
- To evaluate time to progression (TTP) by central radiology review
- To compare TTP on the first and/or the last prior line of systemic therapy (reported) versus TTP on derazantinib by central radiology review (TTP will be calculated from the first date of receiving study drug or prior line of systemic therapy until radiographic disease progression)
- To evaluate ORR, PFS, DCR and TTP by investigator radiology review
- To evaluate the non-confirmed central and local ORR and DCR
- To explore the concordance of local FISH or NGS genomic testing compared to central FISH testing results

The exploratory PK and pharmacodynamic objectives of this study are:

- To evaluate changes in pharmacodynamic biomarkers
- To evaluate ORR, PFS, OS, and DoR by pharmacodynamic biomarkers

- To evaluate the relationship between derazantinib exposure and effectiveness, toxicity and pharmacodynamic biomarkers
- To evaluate PopPK
- To evaluate the urinary excretion of derazantinib, and possibly its metabolites, in a subset of patients

The PK and pharmacodynamic objectives and planned analyses will be further detailed in a separate plan. PK and pharmacodynamic data available at time of database lock and final analysis will be only be listed.

1.3 STUDY ENDPOINTS

1.3.1 Primary Endpoints

Substudy 1
ORR was selected as the primary endpoint of this study. This endpoint is an acceptable surrogate endpoint of clinical benefit in this disease. ORR will be the proportion of patients with confirmed complete responses and partial responses by central radiology review as per RECIST version 1.1.

Substudy 2
PFS 3 will be the proportion of patients who have progression-free survival at 3 months from the first date of receiving study drug as assessed by survival status and central radiology review as per RECIST version 1.1.

1.3.2 Secondary Efficacy Endpoints

Substudy 1 and Substudy 2
- PFS will be calculated from the first date of receiving study drug until radiographic disease progression by central radiology review or death
- OS will be calculated from the first date of receiving study drug until death.
- DoR will be calculated from the first date of documented tumor response to disease progression by central radiology review.
- Changes in HRQOL and symptom response will be evaluated using the EORTC QLQ-C30, QLQ-BIL21, and EQ-5D VAS.

Substudy 2
- ORR will be the proportion of patients with confirmed complete responses and partial responses by central radiology review as per RECIST version 1.1.

1.3.3 Exploratory Efficacy Endpoints

Substudy 1 and Substudy 2
- DCR by central radiology. DCR will be the proportion of patients with confirmed complete or partial responses or stable disease by central radiology review as per RECIST version 1.1.
- TTP will be calculated based on central radiology review from the first date of receiving study drug until radiographic disease progression
- TTP will be assessed for the first and/or the last prior line of systemic therapy (reported) and will be compared to TTP on derazantinib
- ORR, PFS, DCR and TTP by investigator radiology review
- Non-confirmed central and local ORR and DCR
- Concordance of local FISH or NGS genomic testing compared to central FISH testing results

1.3.4 Safety Endpoints

Toxicities will be evaluated using National Cancer Common Terminology Criteria for Adverse Event (NCI CTCAE) version 4.03 criteria.

1.4 SAMPLE SIZES AS PER PROTOCOL, INTERIM ANALYSIS AND DATA MONITORING COMMITTEE

Data Monitoring Committee (DMC)

The DMC was established to ensure the safety of study patients and the validity of study results. The DMC composition and operation was as described in the DMC Charter. The DMC may recommend study termination or continuation based on periodic review of safety and/or efficacy data in this study.

Substudy 1

To prove the drug effect over placebo in ORR in this single-arm pivotal study, the hypothesis was specified as

$$H_0: ORR <= 0.10 \text{ and } H_a: ORR > 0.10.$$

The hypothesis test will be performed at a one-sided 2.5% significance level. A 10% response rate is chosen for the null hypothesis (H_0), which is much higher than the observed placebo rate 7~8% in larger historical trials and publications (Lamarca 2014). The 23% response rate under the alternative hypothesis is estimated for power and sample size calculations.

Approximately 100 patients will be enrolled for this two-stage group sequential study with futility stopping.

An interim analysis for futility was planned to be performed when 40 patients who satisfy mITT criteria had at least one post baseline tumor evaluation. The study was to be terminated for futility if 4 or fewer responses were observed among 40 evaluable patients. If the treatment was ineffective, i.e., ORR=0.1, the probability of futility stopping was 63%. If the true response rate ORR=0.23, the probability of futility stopping was only 3%. The design provided approximately 90% power to reject the null hypothesis at one-sided significance level 0.025, or equivalently to have the lower bound of confidence interval of ORR > 10%. If 5 or more objective responses (OR) were observed and confirmed based on central radiology review prior to enrollment of 40 evaluable patients, the interim analysis was to be performed based on fewer than 40 patients.

The interim analysis was performed on 29 patients satisfying the mITT criteria. At the time of analysis 6 patients (21%) had a confirmed central PR, and futility was rejected per DMC recommendation.

Substudy 2

A Simon's two-stage design will be used in Substudy 2. The null hypothesis (H_0) that the true 3-month rate of PFS 3 is $p_0 \leq 0.45$ will be tested against a one-sided alternative. In the first stage, approximately 15 mITT-evaluable patients will be accrued. If there are 7 or fewer patients with PFS 3 in these 15 patients, then Substudy 2 will be stopped. Otherwise, 28 additional mITT-evaluable patients will be accrued for a total of up to approximately 43. H_0 will be rejected if a PFS 3 is observed in 25 or more of these 43 patients.

The type I error rate is 0.0481, and power is approximately 0.8 when the true PFS 3 rate for derazantinib is $p_1 = 0.65$.

If the required number of patients alive and without disease progression is reached before full enrollment to Stage 1, the decision to transition from Stage 1 to Stage 2 may be taken before Stage 1 is fully enrolled.

If the required number of patients alive and without disease progression is not reached at the time of full enrollment to Stage 1, further enrollment will be suspended to allow for all patients to be evaluated for PFS 3.

The analysis was performed on 14 patients satisfying the mITT criteria. The protocol defined threshold for stage transition and study continuation was achieved with 8 patients out of 14 patients in mITT population meeting the PFS3 criterion.

2 ANALYSIS POPULATIONS

Three analysis populations will be defined as follows:

2.1 SAFETY/INTENT-TO-TREAT POPULATION (ITT)

The Safety/ITT population for Substudy 1 will include all patients with locally or centrally confirmed *FGFR2* fusions/rearrangements who receive any amount of study drug.

The Safety/ITT population for Substudy 2 will include all patients with locally confirmed *FGFR2* mutations or amplifications who receive any amount of study drug.

All data from patients enrolled but not fulfilling the molecular entry criteria (and therefore not included in the Safety/ITT population) will only be listed.

2.2 MODIFIED INTENT-TO-TREAT (MITT) POPULATION

The mITT population will include all patients in the Safety/ITT population, who have at least one post-baseline disease assessment (at least one post-baseline imaging assessment in accordance with RECIST 1.1, or documented clinical progression [every effort should be made to objectively assess radiographic progression]), or reported death during the treatment period

2.3 PER PROTOCOL (PP) POPULATION

The PP population will include all patients in the mITT population who have no important protocol deviations (IPD) which are determined to potentially impact efficacy analyses during the study. Protocol deviations will be identified prior to database lock and final analysis.

3 STATISTICAL AND ANALYTICAL PROCEDURES

3.1 GENERAL STATISTICAL CONSIDERATIONS

3.1.1 Statistical Methods

Summary statistics will be presented as follows. For continuous variables, number of available observations (n), mean, standard deviation (s.d), median, first quartile (Q1), third quartile (Q3), minimum (min), and maximum (max) will be provided; 95% confidence interval for the median will be added as specified. For categorical variables, the number and percentage in each category will be displayed; missing categories will be added accordingly. Percentages will be based on the number of patients in the relevant analysis population.

Time-to-event data will be summarized using the Kaplan-Meier method.

By-patient data listings will be produced for data collected through the study (e.g., case report form, lab). All data listings that contain an evaluation date will contain a relative study day (Study Day). Pre-treatment and on-treatment study days are numbered relative to the day of the first dose of study drug, which is designated as Day 1. The preceding day is Day -1, the day before that is Day -2, etc.

All statistical analyses will be performed in SAS Version 9.4 or higher.

3.1.2 Definition of Baseline

Assessment of change from baseline will include only those patients with both baseline and post-baseline measurements. The last evaluable/non-missing observation/assessment, including unscheduled visits taken before the first dose of study drug will be used as the baseline, unless otherwise specified.

3.1.3 Visit Windows

All data will be tabulated per the evaluation visit as recorded on the electronic case report form, even if the assessment is outside of the visit window. With the exception of the baseline definition, unless otherwise stated in the analysis sections below, unscheduled visits will only be included in data listings.

3.1.4 Handling of Dropouts and Missing Data

In general, unless otherwise specified, missing values will not be imputed for data analysis. Patients whose clinical response is unknown or not reported will be treated as non-responders.

If any laboratory value falls above or below the upper or lower level of quantification, the following rule will be applied for summary statistics: values reported as <XX or ≤XX will be analyzed as XX/2; values reported as >XX or ≥XX will be analyzed as XX (e.g., <0.2 will become 0.1, >0.2 will become 0.2). Reported values will be listed.

All dates are expected to be completed with day, month, and year. For efficacy analyses, if the day/month of death or tumor assessment or last contact date is missing, the 1st of the month/year will be used.

3.2 STATISTICAL METHODS

3.2.1 Disposition of Patients and Analysis Populations

Patient disposition will be presented by substudy and overall. Tabulation will include the following:

- Number (%) of patients treated
- Number (%) of patients who discontinued treatment and primary reason for discontinuation of treatment
- Treatment phase duration (months)
    - Where the treatment phase duration will be calculated as the time from C1D1 to EOT visit. If EOT visit was not performed, then the last study visit prior to EOT will be analyzed.
- Number (%) of patients who performed the 30-day Safety Follow-up visit
- Number (%) of patients who entered Survival Follow-up, their last outcome at time of analysis (alive, continuing in follow-up, dead, lost to follow-up, alive, withdrawal of consent from study follow-up)
- Study duration (months)
    - Where the study duration will be calculated as the time from C1D1 to last survival follow-up date, or death. If last survival follow-up date, or death is missing, then the last date of contact will be analyzed.

The table will be based on the Safety/ITT Population.

The number and percentage of patients in each analysis populations (Safety/ITT, mITT, PP), including reasons for exclusion from each analysis population will be presented for all patients enrolled i.e. patients signed with informed consent.

The number of patients by country and site will be presented by the Safety/ITT population.

The following by-patient listings will be presented by substudy,

- Patient eligibility with inclusion and exclusion criteria details
- Patient enrollment
- Patient disposition
- Analysis populations
- IPDs

3.2.2 Demographics and Baseline Characteristics

Demographic and baseline characteristics will be summarized by the Safety/ITT, mITT and PP populations.

The following variables will be listed and summarized in the demographic and baseline characteristic summary table using both continuous and categorical descriptive statistics.

- gender,
- race (White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Other, not allowed to record per local regulation, unknown)
- ethnicity (Hispanic or Latino, not Hispanic or Latino, not reported),
- age (years) at baseline,
- baseline Eastern Cooperative Oncology Group (ECOG),
- baseline vital signs: height (cm), weight (kg), body mass index (BMI) (kg/m^2) – calculated as weight (kg) / height $(m)^2$
- tumor's genomic status (*FGFR2* fusion for Substudy 1, and *FGFR2* mutations or amplification or

Substudy 2)

Concordance of local FISH or NGS genomic testing compared to central FISH testing results will be summarized descriptively.

By patient listing of demographics, pregnancy test, tumor's genomic status will be produced,

3.2.3 Medical History (Non-Cancer Related)

Medical history will be coded using Medical Dictionary for Regulatory Activities (MedDRA) version 23.0 or later and will be summarized by system organ class (SOC) and preferred term (PT) for the Safety/ITT population. Patients will be counted only once for each SOC or PT in the event that they have multiple records of the same SOC or PT in the database. All medical history data will be listed.

3.2.4 Cancer-Related Medical History

Cancer-related medical history will be summarized for the Safety/ITT population. Descriptive statistics on cancer-related medical history will be presented for the following:

- Cancer history for iCCA
- Prior systemic cancer therapies
- Prior iCCA local therapies
- Prior iCCA cancer surgery/procedure
- Prior liver resection:
- Prior iCCA radiation therapies:

All cancer-related medical history data will be listed.

3.2.5 Prior and Concomitant Medications, and Non-Drug Treatments/Procedures

Prior and concomitant medications will be summarized for the Safety/ITT population. All prior and concomitant medications will be included in by-patient data listings.

Non-drug treatments/procedures will only be listed for the Safety/ITT population.

Prior medication will be defined as any medication taken and stopped prior to the first dose of study drug. Concomitant medication will be defined as any medication ongoing at time of first dose of study drug or taken after the first dose of study drug.

Medications missing both start and stop dates, or having a start date prior to the first dose of study drug and missing the stop date, or having a stop date on or after the last dose of study drug and missing start date will be counted as concomitant.

For partial dates, the following approach will be taken:

• If the start day is missing but the start month and year are complete, a medication will be excluded as being concomitant only if the start month/year is before the month/year of study drug administration and if the stop date (either full date, month and year if missing day, or year if missing month and day) is before study drug administration.

• If the start day and month are missing but the start year is complete, a medication or procedure will be excluded as concomitant only if the start year is before the year of study drug administration and if the stop date (either: full date, month and year if missing day, or year if missing month and day) is before study drug administration.

Prior and concomitant medications will be coded using the WHO Drug Dictionary (WHODRUG Global B3 Sep2019 or above), and patient incidence will be tabulated by Anatomic Therapeutic Class (ATC) level 2, level 4 and PT. Patients will be counted only once for each ATC or PT in the event that they have multiple records of the same ATC or PT in the database.

3.2.6 Study Drug Exposure and Compliance

Study drug exposure and compliance will be presented using the Safety/ITT population.

Exposure to study drug will be presented cumulatively.

The number of cycles of study drug received is defined as the total number of cycles where the patient has received at least 1 dose of study drug in each cycle, and excluding any cycle(s) where the patient has not been dosed. The study drug exposure is recorded for each cycle where the number of capsules dispensed and number of capsules returned will be entered; where the number of capsules returned is equal to the number of capsules dispensed, then this indicates that the patients were not dosed at this specific cycle.

Duration of study drug is defined as: (date of last dose – date of first dose) + 1. If the date of last dose is missing then this will be imputed as the date of the last study drug dispensed.

Dose intensity (%) is defined as the percentage of the study drug a patient has taken compared to the amount of study drug the patient was meant to take. This will be calculated as:

Total cumulative dose received (mg) / total cumulative dose planned (mg) *100

Total cumulative dose received (mg) is defined as the sum of all dose (mg) as recorded from the study drug administration eCRF page using each entry (date of last dose –date of first dose) + 1 * dose in mg. If the date of last dose is missing then this will be imputed with the date of the last study drug dispensed.

Similarly, the total cumulative dose planned (mg) is defined as the sum of all planned dose (mg) as recorded from the study drug administration eCRF page using each entry (date of last dose –date of first dose) + 1 * 300 mg. If the date of last dose is missing then this will be imputed with the date of the last study drug dispensed.

Treatment compliance is defined as the percent of actual amount of drug taken relative to the planned amount. It is calculated according to the following formula for the entire study:

% compliance = (# of capsules dispensed - # of capsules returned)/{(# of days in the dosing interval)*(# capsules per daily dose)}*100

where the # of days in the dosing interval excludes any days that the patient was instructed to hold dosing due to an AE.

Treatment compliance will also be summarized categorically, considering cut-offs of 80% and 120% for compliance.

The number of cycles of study drug received, total duration of study drug (weeks), dose density (%), total cumulative dose received (mg) and treatment compliance (%) will be summarized using continuous statistics. Treatment compliance (%), the number of patients with at least 1 dose change/interruption during

the duration of the study, and the reasons for dose change/interruption will be summarized as categorical variables.

Study drug administration, drug accountability, exposure and compliance will be presented by patient listings.

3.2.7 Analyses of Efficacy Variables

All efficacy data will be summarized by substudy as detailed below, and listed.

For Substudy 1, the Safety/ITT population will be used to analyze the efficacy data, with the mITT and PP population used for sensitivity analyses. For Substudy 2, the mITT population will be used to analyze the efficacy data, with the ITT and PP population used for the sensitivity analyses.

Assessments by Blinded Independent Central Review (BICR) consider two readers. Should there be disagreement between these two readers, an adjudicator decides which reader is correct. As such, all analyses which consider BICR assessments will be based upon the first reader by default, unless the second reader is selected following adjudication. BICR and local assessment will be analyzed and listed.

Waterfall plots, spider plots and swimmer plots will be produced to accompany efficacy analyses.

3.2.7.1 Analysis of Primary Efficacy Variable

Substudy 1

The primary efficacy endpoint for Substudy 1 will be ORR, defined as the achievement of confirmed CR or PR using RECIST v1.1(Eisenhauer 2009) as assessed by BICR.

Per protocol inclusion criteria, only patients with measurable disease by RECIST version 1.1 criteria should be enrolled. If a patient without measurable disease is enrolled, the ITT principle requires including these patients in the analyses. Hence, analyses will be based on patients with either measurable or non-measurable disease. For this purpose, non-CR/non-PD in patients with non-measurable disease will be considered equivalent to SD in patients with measurable disease, and the same rules will be applied.

Classification of best confirmed response is done according to the following hierarchy and rules:

1. CR: requires two consecutive CR response assessments a minimum of four weeks apart
2. PR: requires two consecutive PR response assessments OR a PR response followed by a consecutive CR response assessment (a minimum of four weeks apart)
3. SD: requires only one SD response assessment (provided minimum criteria for SD duration met)
4. Non-CR/Non-PD (non-measurable disease patients only): requires only one Non-CR/Non-PD response assessment (provided minimum criteria for Non-CR/Non-PD duration met)
5. PD: requires only one PD response assessment

The minimum duration for SD and Non-CR/Non-PD is defined as at least 6 weeks i.e. 42 days, where the duration is calculated from the start of treatment until the criteria of SD is met.

The classification as per RECIST 1.1 (Eisenhauer 2009) is summarized in Table 1 below.

Table 1: Best overall response when confirmation of CR and PR required.

Overall response		BEST overall response
First time point	**Subsequent time point**	
CR	CR	CR
CR	PR	SD, PD or PR*
CR	SD	SD provided minimum criteria for SD duration met, otherwise PD
CR	PD	SD provided minimum criteria for SD duration met, otherwise PD
CR	NE	SD provided minimum criteria for SD duration met, otherwise NE
PR	CR	PR
PR	PR	PR
PR	SD	SD
PR	PD	SD provided minimum criteria for SD duration met, otherwise PD
PR	NE	SD provided minimum criteria for SD duration met, otherwise NE
NE	NE	NE

CR = complete response, PR = partial response, SD = stable disease, PD = progressive disease, and NE = inevaluable.
* If a CR is truly met at first time point, then any disease seen at a subsequent time point, even disease meeting PR criteria relative to baseline, makes the disease PD at that point (since disease must have reappeared after CR). Best response would depend on whether minimum duration for SD was met. However, sometimes 'CR' may be claimed when subsequent scans suggest small lesions were likely still present and in fact the patient had PR, not CR at the first time point. Under these circumstances, the original CR should be changed to PR and the best response is PR.
For patients with non-measurable disease, SD should be replaced by non-CR/Non-PD

If the best overall response has any special cases or difficulty from unexpected data, then these may be checked and discussed with the study team to decide which best overall response is best suited.

As specified in RECIST 1.1, repeated 'NE' time point assessments complicate best response determination. Should a CR/PR be followed by one or more NE response and then a CR/PR, then confirmed response would be achieved (as an example PR-NE-PR will be considered as a confirmed response).

Patients without efficacy assessments are considered non-responders, and included in the analysis.

Hypothesis testing (null hypothesis [H_0] versus alternative hypothesis [H_A]) were performed at the interim and will be performed in the final analyses as specified below:

Null Hypothesis H_0: ORR <= 0.10 and
Alternative Hypothesis H_A: ORR > 0.10

The hypothesis testing will be performed using exact test for binomial proportion at a one-sided 2.5% significance level. Point estimates and exact 2-sided 95% confidence intervals (Clopper-Pearson) will be provided. If the p-value is ≤ to 2.5% (one-sided), or equivalently the lower limit of the confidence interval is > 10%, the null hypothesis will be rejected.

Considering the number of patients per site planned in this study, the site effect and treatment by site interaction will not be evaluated.

Substudy 2

The primary efficacy endpoint for Substudy 2 will be the proportion of patients who have PFS at 3 months from the first date of receiving study drug as assessed by survival status or central radiology review RECIST 1.1.

PFS will be calculated as the time from first dose of study drug until disease progression as assessed by BICR, or death from any cause, whichever occurs first. Patients who either have no baseline tumor evaluation or have no post-baseline tumor evaluation will be censored at date of first dose. Patients who discontinue treatment due to reasons other than disease progression by BICR or death will be censored in the PFS analyses as the date of their last tumor evaluation prior to EOT. Patients who progress or die after missing ≥2 consecutive scheduled tumor assessments will be censored at the date of the last tumor evaluation prior to progression or death. Any disease progression as assessed by BICR, or death occurring after end of treatment visit are not taken into account as a PFS event. For missing end of treatment visit date, a +7 day window compared to the maximum of last dose date, last on treatment visit date will be applied to create a tentative end of treatment visit date and a cut-off for death occurring after this date.

The rules for censoring PFS are summarized in the table below:

Table 2: Censoring scheme for PFS

Situation	Date of progression or censoring	Outcome
No baseline tumor assessments	Date of first dose	Censored
No post-baseline tumor evaluation	Date of first dose	Censored
Documented radiological progression	Date of radiological assessment of measured lesions up to EOT	Progressed
No radiological progression	Date of last radiological assessment of measured lesions up to EOT	Censored
Treatment discontinuation for reasons other than radiological disease progression by BICR or death	Date of last radiological assessment of measured lesions up to EOT	Censored
Death before first PD assessment	Date of death	Progressed
Death between adequate assessment visits	Date of death	Progressed
Death or radiological progression after more than two missed tumor assessment or after end of treatment	Date of last radiological assessment of measured lesions prior to progression or death	Censored

The hypothesis being tested is:

Null Hypothesis H_0: PFS 3 ≤ 0.45 and
Alternative Hypothesis H_A: PFS 3 > 0.45

Where PFS 3 is the proportion of patients with a PFS > 3 months (i.e. > 91 days).

A Simon's two-stage design will be used, with approximately 15 patients in Stage 1, and an additional 28 patients if the study proceeds to Stage 2. H_0 will be rejected if a PFS 3 is observed in 25 or more of these 43 patients. The type I error rate is 0.0481, and power is approximately 0.8 when the true PFS 3 rate for derazantinib is p1 = 0.65.

Point estimates, exact 2-sided 95% confidence intervals (Clopper-Pearson) and associated p-value will be provided.

PFS will be analyzed using Kaplan-Meier methodology (Kaplan & Meier 1958). The median duration of PFS will be presented along with the standard error and 2-sided 95% CI. If they are calculable, the 25th and 75th percentiles and the 2-sided 95% CIs around the percentiles will be presented. Kaplan-Meier estimates at 3, 6, 9 and 12 months will also be presented. The Kaplan-Meier survival curves will also be presented.

Additional PFS sensitivity analyses which consider alternative censoring methods as outlined in FDA guidelines (FDA 2018, FDA 2015) will be conducted in the same manner, as described above. These will be performed for the Safety/ITT and mITT populations:

- Sensitivity analysis 1 in Table 3 uses a conservative approach by assigning the date of discontinuation, or missed tumor assessment followed by PD or death as an event date.

- Sensitivity analysis 2 in Table 4 includes clinical progression as an event.

- Sensitivity analysis 3 in Table 5 evaluates PFS including post EOT radiological assessments.

Table 3: Censoring scheme for PFS - sensitivity analysis 1 (any change considered as progression event)

Situation	Date of progression or censoring	Outcome
No baseline tumor assessments	Date of first dose	Censored
No post-baseline tumor evaluation	Date of first dose	Censored
Documented radiological progression between scheduled visits	Date of radiological assessment of measured lesions	Progressed
No radiological progression	Date of last radiological assessment of measured lesions	Censored
Treatment discontinuation for reasons other than radiological disease progression by BICR or death	Date of discontinuation	Progressed
Death before first PD assessment	Date of death	Progressed
Death between adequate assessment visits	Date of death	Progressed
Death or radiological progression after more than two missed tumor assessment	Date of first missed tumor assessment	Progressed

Table 4: Censoring scheme for PFS - sensitivity analysis 2 (includes clinical progression)

Situation	Date of progression or censoring	Outcome
No baseline tumor assessments	Date of first dose	Censored
No post-baseline tumor evaluation	Date of first dose	Censored
Documented radiological progression	Date of radiological assessment of measured lesions up to EOT	Progressed
No radiological progression, no clinical progression	Date of radiological assessment of measured lesions up to EOT	Censored
Clinical progression	Date of clinical progression	Progressed
Treatment discontinuation for reasons other than radiological disease progression by BICR, clinical progression or death	Date of radiological assessment of measured lesions up to EOT	Censored
Death before first PD assessment	Date of death	Progressed
Death between adequate assessment visits	Date of death	Progressed
Death or radiological progression or clinical progression after more than two missed tumor assessment or after end of treatment	Date of last radiological assessment of measured lesions prior to progression or clinical progression or death	Censored

Table 5: Censoring scheme for PFS - sensitivity analysis 3 (including post EOT assessments)

Situation	Date of progression or censoring	Outcome
No baseline tumor assessments	Date of first dose	Censored
No post-baseline tumor evaluation	Date of first dose	Censored
Documented radiological progression	Date of radiological assessment of measured lesions	Progressed
No progression, or new anticancer treatment started	Earliest date of: - last radiological assessment of measured lesions, including post EOT - new anticancer treatment	Censored
Death before first PD assessment	Date of death	Progressed
Death between adequate assessment visits	Date of death	Progressed
Death or progression after more than two missed tumor assessment	Date of last radiological assessment of measured lesions prior to progression or death	Censored

3.2.7.2 Analyses of Secondary Efficacy Variables

3.2.7.2.1 PFS (Substudy 1 and Substudy 2)

PFS as assessed by BICR will be analyzed as detailed in Section 3.2.7.1 for primary efficacy analysis for Substudy 2.

3.2.7.2.2 DoR (Substudy 1 and Substudy 2)

DoR (months) will be calculated from the first date of documented objective tumor response (confirmed CR or PR) to disease progression as assessed by BICR or death. If a patient is discontinued or is lost to follow-up with no documentation of progressive disease, DoR is defined as the time from the date of the first documentation of OR to the date of the last tumor assessment as a censored value.

DoR will be derived only for patients who have the best overall response of CR or PR.

DoR will be analyzed using Kaplan-Meier methodology (Kaplan & Meier 1958). The median DoR will be presented along with the standard error and 2-sided 95% CI. If they are calculable, the 25th and 75th percentiles and the 2-sided 95% CIs around the percentiles will be presented. Kaplan-Meier estimates at 3, 6, 9 and 12 months will also be presented. The Kaplan-Meier survival curves will also be presented.

The date of first study drug dose, documented objective tumor response, disease progression, death, last tumor assessment, last contact and DoR days will be reported in a by-patient listing.

3.2.7.2.3 OS (Substudy 1 and Substudy 2)

OS time (months) will be calculated from the first date of receiving study drug until death from any cause. Any patient without a date of death in the database at the time the survival analyses are performed will be censored at the time of their last study contact. OS will be performed for the Safety/ITT population.

OS will be analyzed using Kaplan-Meier methodology (Kaplan & Meier 1958). The median duration of OS will be presented along with the standard error and 2-sided 95% CI. If they are calculable, the 25th and 75th percentiles and the 2-sided 95% CIs around the percentiles will be presented. Kaplan-Meier estimates at 3, 6, 9 and 12 months will also be presented. The Kaplan-Meier survival curves will also be presented.

The date of first study drug dose, death, last study contact and OS days will be reported in a by-patient listing.

3.2.7.2.4 ORR (Substudy 2)

ORR will be considered a secondary efficacy endpoint in Substudy 2 and analyzed in the same way as stated in Section 3.2.7.1.

3.2.7.2.5 PRO Assessments (Substudy 1 and Substudy 2)

For each assessment the minimally important difference (MID) in a score at C3D1 and C5D1 will be determined following the recent protocol by the EORTC Quality of Life Group (Musoro 2018). Patients with complete data will be categorized into five mutually exclusive clinical change groups (CCG) reflecting the five possible levels of change. For each pair of timepoints (Change in summary score from baseline at

C3D1 and Change in summary score from C3D1 at C5D1), a patient can thus belong to only one CCG category. Two methods for determining the MID will be explored:

1. **Mean Change Method**
   For a given mean absolute change in PRO assessment scale summary score: the MID for improvement is equal to the mean summary score of the 'small positive change' CCG (somewhat better now than 6 weeks ago) and the MID for deterioration is equal to the mean summary score of the 'small negative change' CCG (somewhat worse now than 6 weeks ago). The mean summary score of the 'small change' CCGs and that of the 'no change' CCG will be compared and reported. If the mean summary score for 'no change' CCG is similar to any of the two 'small change' CCGs, the estimated MID may be considered doubtful. No a priori rule will be established to determine if mean summary scores are sufficiently similar to determine MID to be doubtful.

2. **Linear Regression**
   For a given absolute change in PRO assessment scale summary score: the estimate of the numerical change in summary score that is associated with the transition between adjacent CCG categories will be determined using linear regression. Separate models will be fitted for improving and deteriorating scores based on the anchor. The outcome variable is the PRO assessment summary score, and the covariate is a binary anchor variable; coded as 'no change'=0 and 'small positive change'=1 for model on improvement (Other CCG categories are excluded from the linear regression model), and 'no change'=0 and 'small negative change'=1 for model on deterioration (Other CCG categories are excluded from the linear regression model. The resulting β's (i.e. slope parameters) correspond to the MIDs for improvement and deterioration respectively. No other covariates will be included in these models.

   PRO analysis will be performed for the Safety/ITT population.

3.2.7.2.5.1 G-SET/HTI

G-SET/HTI is a single item, and will be used as an external anchor to determine the minimal important difference of EQ-5D-VAS, EORTC QLQ-C30, and QLQ-BIL21 scales. The G-SET/HTI is a patient-rated change in health between two time periods using a five-point ordinal scale (1=much better now than 8 weeks ago; 2=somewhat better now than 8 weeks ago; 3=about the same as 8 weeks ago; 4=somewhat worse now than 8 weeks ago; 5=much worse now than 8 weeks ago). G-SET/HTI will be summarized at Cycle 3, Day 1, and Cycle 5, Day 1 by categorical and continuous descriptive statistics.

3.2.7.2.5.2 EORTC QLQ-C30

EORTC QLQ-C30 was developed to assess the quality of life of cancer patients and is the most widely used cancer-specific HRQoL instrument. It contains 30 items and measures five functional scales (physical, role, cognitive, emotional, and social); three symptom scales (fatigue, pain, and nausea and vomiting); six single symptom items (dyspnea, sleep disturbance, appetite loss, constipation, diarrhea and financial impact) and a global health and quality-of-life scale. The global health and quality of life items uses a 7-point scale scoring from 1 (very poor) to 7 (excellent). The other items are scored on a 4-point scale from 1 (not at all) to 4 (very much). A high score for a functional scale represents a high/healthy level of functioning. A high score for the global health status represents a high QoL. However, a high score for a symptom scale/item represents a high level of symptomatology/problems. The scoring method is outlined in Table 6 and the following paragraphs.

Table 6: Scoring the QLQ-C30 version 3.0

	Scale	Number of items	Item range	Version 3.0 Item numbers	Function scales
Global health Status /QoL	QL	2	6	29, 30	
Functional Scales					
Physical functioning	PF	5	3	1 to 5	F
Role functioning	RF	2	3	6,7	F
Emotional functioning	EF	4	3	21 to 24	F
Cognitive functioning	CF	2	3	20, 25	F
Social function	SF	2	3	26, 27	F
Symptom scales/items					
Fatigue	FA	3	3	10, 12, 18	
Nausea and vomiting	NV	2	3	14,15	
Pain	PA	2	3	9, 19	
Dyspnea	DY	1	3	8	
Insomnia	SL	1	3	11	
Appetite loss	AP	1	3	13	
Constipation	CO	1	3	16	
Diarrhea	DI	1	3	17	
Financial difficulties	FI	1	3	28	

Raw score (RS) is calculated as the average of item score when at least half of the items are not missing. A linear transformation to 0–100 will then be applied to get the score (S) using the item range provided in Table 6.

For functional scales S = (1-(RS-1)/range) x 100

For symptom scales/items and global health status S = ((RS-1)/range) x 100

Each scale will be summarized at each visit as a continuous variable using descriptive statistics per substudy for the total scores, change from baseline and percent change from baseline for all visits where the EORTC QLQ-C30 is collected. In addition, at C5D1, the change from C3D1 at C5D1 and the percentage change at C5D1 from C3D1 will also be reported. For each of the functional scales, the minimally important change from baseline at C3D1 and change from C3D1 at C5D1 will be determined using the two methods detailed in Section 3.2.7.2.5.

3.2.7.2.5.3 EORTC QLQ-BIL21

The EORTC QLQ-BIL21 comprises five multi-item measures of cholangiocarcinoma and gallbladder cancer-associated symptoms (eating symptoms, jaundice, tiredness, pain symptoms, anxiety), and three single items on treatment side effects, difficulties with drainage bags/tubes, and concerns regarding weight loss. The response is a four-point Likert scale.

Table 7: Scoring the QLQ-BIL21

	Number of items	Version 3.0 Item numbers
Eating	4	31, 32, 33, 34
Jaundice	3	35, 36, 37
Tiredness	3	38, 39, 40
Pain	4	41, 42, 43, 44
Anxiety	4	45, 46, 47, 48
Treatment Side Effects	1	49
Drainage Bag/Tube Difficulties	1	50
Weight Loss Concerns	1	51

Raw score (RS) is calculated as the average of item score when at least half of the items are not missing. A linear transformation to 0–100 will then be applied to get the score (S) using the item range provided in Table 7.

For functional scales S = (1-(RS-1)/range) x 100

For symptom scales/items and global health status S = ((RS-1)/range) x 100

Each scale will be summarized at each visit as a continuous variable using descriptive statistics per substudy for the total scores, change from baseline and percent change from baseline for all visits where the EORTC QLQ-BIL21 is collected. In addition, at C5D1, the change from C3D1 at C5D1 and the percentage change at C5D1 from C3D1 will also be reported. For each of the functional scales, the minimally important change from baseline at C3D1 and change from C3D1 at C5D1 will be determined using the two methods detailed in Section 3.2.7.2.5.

3.2.7.2.5.4 EQ-5D (5L)

The EQ-5D (5L) is a standardized instrument for use as a measure of health outcome. The five health state dimensions in this instrument include the following: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension is rated on a five-point scale from 1 (no problem) to 5 (extreme problem). The EQ-5D (5L) also includes a graded (0 to 100) VAS on which the patient rates his or her general state of health at the time of the assessment.

The 5 questions will be summarized descriptively as categorical variables for all visits where the EQ-5D (5L) is collected. The VAS will be summarized as a continuous variable using descriptive statistics for the total scores, change from baseline and percent change from baseline for all visits where the EQ-5D (5L) is collected. In addition, at C5D1, the change from C3D1 at C5D1 and the percentage change at C5D1 from C3D1 will also be reported. The minimally important change from baseline at C3D1 and change from C3D1 at C5D1 of the EQ-5D (5L) VAS will be determined using the two methods detailed in Section 3.2.7.2.5. Missing VAS scores will not be included in summary statistics and only patients with both an observation at the specified visit and at baseline will be included in change from baseline summaries.

3.2.7.3 Exploratory Analyses

3.2.7.3.1 Disease Control Rate (Substudy 1 and Substudy 2)

Disease control rate (DCR) is defined as the achievement of confirmed CR or PR or SD (or Non-CR/Non-PD for patients with non-measurable disease) using RECIST 1.1 as assessed by BICR. Point estimates and exact 2-sided 95% CIs (Clopper-Pearson) will be provided by the Safety/ITT Population.

DCR by investigator assessment will also be analyzed for the Safety/ITT Population.

3.2.7.3.2 Investigator Assessed Best Overall Response (Substudy 1 and 2)

In the primary analysis, tumor response will be based upon measurements evaluation by the BICR. However, an initial assessment of tumor response will be made by investigators at site. As an exploratory measure, analyses of best overall response will be conducted considering investigator assessment only, these will be summarized by the Safety/ITT, mITT and PP population.

3.2.7.3.3 Investigator Assessed PFS (Substudy 1 and 2)

PFS as assessed by the investigators will be analyzed similarly as detailed in Section 3.2.7.2.1 for the Safety/ITT and mITT Population.

3.2.7.3.4 TTP (Substudy 1 and 2)

TTP will be calculated as follows:

a. TTP (on derazantinib): TTP will be calculated from the first date of receiving study drug until first radiographic disease progression per BICR.
b. TTP (on the first prior line of systemic therapy): TTP will be calculated from the date of first prior line of systemic therapy until date of radiographic disease progression reported for that regimen
c. TTP (on the last prior line of systemic therapy): TTP will be calculated from the date of last prior line of systemic therapy until date of first radiographic disease progression reported for that regimen

Missing dates for the prior line of systemic therapy will be imputed as detailed in Section 3.1.4. The dates for the prior lines will be sorted by ascending dates to determine the first and last lines of systemic therapy.

Censoring of the TTP variables are summarized in Table 8, 9 and 10 below:

Table 8: Censoring scheme for TTP (on derazantinib)

Situation	Date of progression or censoring	Outcome
No baseline tumor assessments	Date of first dose	Censored
No post-baseline tumor evaluation	Date of first dose	Censored
Radiological progression on or prior to EOT	Date of radiological assessment of measured lesions prior to EOT	Progressed
No radiological progression	Date of last radiological assessment of measured lesions prior to EOT	Censored
Treatment discontinuation for reasons other than radiological	Date of last radiological assessment of measured lesions prior to EOT	Censored

disease progression by BICR		
Death	Date of last radiological assessment of measured lesions prior to EOT	Censored
Radiological progression after more than two missed tumor assessment or after end of treatment	Date of last radiological assessment of measured lesions prior to radiological progression, or EOT	Censored

Table 9: Censoring scheme for TTP (first line)

Situation	Date of progression or censoring	Outcome
No radiological progression reported for first regimen	Date of last dose of treatment for the regimen	Censored

Table 10: Censoring scheme for TTP (last line)

Situation	Date of progression or censoring	Outcome
No radiological progression reported for last progression	Date of last dose of treatment for the regimen	Censored

This will be analyzed using Kaplan-Meier methodology (Kaplan & Meier 1958). The median duration of TTP will be presented along with the standard error and 2-sided 95% CI. If they are calculable, the 25th and 75th percentiles and the 2-sided 95% CIs around the percentiles will be presented. Kaplan-Meier estimates at 3, 6, 9 and 12 months will also be presented. The Kaplan-Meier survival curves will also be presented.

TTP by line of prior systemic therapy will also be performed similarly.

TTP assessed on the first and the last prior line of systemic therapy (reported) and TTP on derazantinib will be presented by a Kaplan-Meier survival curve.

3.2.8 Analyses of Safety Data

Safety analyses will be performed for the Safety/ITT population.

3.2.8.1 Adverse Events

AEs will be coded using the MedDRA and assigned the grades (severities) based on NCI CTCAE, Version 4.03. For AEs not listed in the NCI CTCAE version 4.03, a similar grading system should be used as follows:

- Grade 1: Mild AE
- Grade 2: Moderate AE
- Grade 3: Severe AE
- Grade 4: Life-threatening or disabling AE
- Grade 5: Death

A treatment-emergent adverse event (TEAE) is defined as an AE with start date on or after the date of first dose and up to 30 days after last dose of study treatment (or until the start of a new anticancer therapy, whichever occurs earlier).

An overall TEAE incidence summary table will be produced and will include the number and percentage of patients with any TEAE, with any TEAE assessed by the Investigator as related to drug (related/not related), and with any serious TEAEs, drug-related serious TEAEs, TEAEs leading to permanent study drug discontinuation, drug-related TEAEs leading to permanent study drug discontinuation, TEAEs leading to study drug dose reduction or interruption, TEAEs leading to study drug dose reduction, TEAEs leading to study drug interruption, drug-related TEAEs leading to study drug dose reduction or interruption, drug-related TEAEs leading to study drug dose reduction, drug-related TEAEs leading to study drug interruption TEAEs leading to death, drug-related TEAEs leading to death will be summarized. The number of events will also be displayed.

For all event types listed above in the overall AE incidence summary table, tabulations by SOC and PT will also be produced. These will be presented by SOC and PT, sorted by decreasing frequency. Furthermore, TEAE, and drug-related TEAEs will also be summarized by SOC, PT, and maximum CTCAE grade, and sorted by decreasing frequency. Patients with more than one AE within a particular SOC are counted only once for that SOC. Similarly, patients with more than one AE within a particular PT are counted only once for that PT. For summaries by maximum CTCAE grade, patients with multiple AEs within a particular SOC or PT will be counted under the category of their most severe AE within that SOC or PT. AEs with missing CTCAE grade will be included in the missing counts.

The following TEAEs and drug-related TEAEs of special interest will also be presented by SOC and PT and sorted by decreasing frequency: Transaminase elevations, hyperphosphatemia, blood creatinine increased, nail toxicity, dry eye, corneal, and specific eye TEAEs (keratitis, retinal events, asthenia/fatigue, dry eye/xerophthalmia).

No formal hypothesis-testing analysis of AE incidence rates will be performed.

TEAEs will be listed in patient data listings. By-patient listings also will be provided for the following: patient deaths, serious adverse events (SAEs), and AEs leading to withdrawal.

3.2.8.2 Laboratory Variables

Local laboratory results will only be listed.

Central laboratory results will be summarized with descriptive statistics at each scheduled time and for the change from baseline. Repeated or unscheduled tests will not be summarized for each scheduled visit (with the exception of baseline which may include unscheduled visits). Baseline values and values at each visit for each laboratory parameter collected be descriptively summarized and plotted over time using a boxplot.
Laboratory data will also be summarized using shift tables where appropriate. Each patient's continuous laboratory safety parameter values will be flagged as "low", "normal", "high" or "missing" relative to the normal ranges. Each patient's categorical laboratory safety parameter values will be flagged as "abnormal" or "normal". This categorical data will be summarized in shift tables comparing the minimum post-baseline value, maximum post-baseline value and all other relevant post-baseline visits with those at the baseline visit. Note that minimum post-baseline will only be displayed in the instances where there is a range with lower limit of normal (LLN) >0. For parameters that have gradings for both low and high values, low and high will be summarized separately.

In the shifts of low values, any high values will be classified as normal and the worst low grade will be summarized. And in the shifts of high values, any low values will be classified as normal and the worst high grade will be summarized.

In addition, CTCAE grade and hyperphosphatemia grade will be summarized. Shift tables of grade from baseline to maximum or minimum on-treatment as applicable will be produced.

Since hyperphosphatemia is not defined by CTCAE v. 4.03, as per protocol hyperphosphatemia is defined as:
Grade 1: > upper limit of normal (ULN) to < 7.0 mg/dL (<2.26 mmol/L)
Grade 2: Non-invasive intervention required (e.g., withhold drug or modify dose) or between 7.0 – 9.0 mg/dL (2.26 – 2.90 mmol/L)
Grade 3: Severe or medically significant, but not immediately life threatening, or > 9.0 – 10.0 mg/dL (> 2.90 – 3.23 mmol/L)
Grade 4: Life-threatening consequences, urgent intervention indicated e.g., dialysis, or > 10.0 mg/dL (> 3.23 mmol/L)

Central and local laboratory results will be included in by-patient listings and will include CTCAE grade and hyperphosphatemia grade at each visit, as well as change from baseline at each visit.

Hy's law

Patients who have elevated alanine aminotransferase (ALT), aspartate aminotransferase (AST), and total bilirubin post baseline will be summarized descriptively as follows and Hy's law cases identified.

- ALT and AST: > 3x ULN, > 5x ULN, > 10x ULN, > 20x ULN
- Total bilirubin >2x ULN
- ALT or AST: > 3x ULN and Total bilirubin >2x ULN

Plots of ALT and AST vs. Total bilirubin by module will also be produced with reference lines at 3×ULN for ALT, AST, and 2×ULN for total bilirubin. In each plot, peak total bilirubin x ULN will be on the vertical axis and peak ALT or AST x ULN will be on the horizontal axis.

3.2.8.3 Vital Sign Variables

Values and change from baseline for vital signs, including systolic blood pressure, diastolic blood pressure, heart rate, respiratory rate, body temperature, and weight will be summarized by scheduled time of evaluation. Repeated or unscheduled tests will not be summarized for each scheduled visit.

By patient listings for all vital signs will also be presented.

3.2.8.4 Electrocardiogram (ECG)

The actual value and change from baseline to each visit and at end of study will be descriptively summarized by visit for ECG parameters (e.g., heart rate, RR interval, PR interval, QRS duration, QT interval, QTcF interval and QTcB interval). For all patients, a standard, triplicate, 12-lead ECG must be performed at all study visits using the pre-programmed device provided by the Sponsor. Summary tables will consider the mean of the triplicate measurements at each visit as the single reported value.

A categorical analysis of QTcF interval and QTcB interval correction will also be presented, based on the number and percentage of patients meeting or exceeding specific thresholds for absolute QTcF/QTcB interval prolongation and change from baseline in QTcF/QTcB interval. In this analysis, the mean of the triplicate measurements will not be used; instead, the maximum at each timepoint will be considered.

For absolute QTcF/QTcB interval prolongation, the number and percentage of patients within the following thresholds will be summarized at each visit. The worst (largest) value will also be summarized.

- Interval ≤450 ms
- Interval >450 ms and ≤480 ms
- Interval >480 ms and ≤500 ms
- Interval >500 ms

The change from baseline in QTcF/QTcB interval to the worst (largest) post-baseline observation and at each study visit will also be summarized, considering the following thresholds:

- Interval increased from baseline ≤30 ms
- Interval increased from baseline >30ms and ≤60 ms
- Interval increased from baseline >60 ms

For the analysis of change from baseline, the mean of the triplicate measurements will be used for baseline, but the maximum of triplicate measurements will be used for post-baseline values.

Local and central ECG data will be presented separately in the summary tables, and will be presented by timepoint. All ECG data will be provided in a by-patient listing.

3.2.8.5 ECOG

ECOG performance status will be summarized in a shift table, which will summarize values at baseline, each visit and the maximum values during the study.

All ECOG performance assessments will be provided in a by-patient listing.

3.2.8.6 Other Safety Variables

By patient listings for physical examination, eye examination, and pregnancy test will be presented.

4 CHANGES FROM PLANNED ANALYSIS

- The protocol indicates the primary analysis population is restricted to FISH performed by a central laboratory. However, this was not the case and therefore reference to central FISH has been removed

- Safety/ITT Population has been updated from "All patients who receive any amount of study drug" to 2 separate definitions for each substudy to take into consideration the substudy that each patient was enrolled in:

    - The Safety/ITT population for Substudy 1 will include all patients with locally or centrally confirmed *FGFR2* fusions/rearrangements who receive any amount of study drug.

    - The Safety/ITT population for Substudy 2 will include all patients with locally confirmed *FGFR2* mutations or amplifications who receive any amount of study drug.

- The mITT population definition was updated from "All patients who receive any amount of study drug and have at least one post-baseline disease assessment." to "All patients in the Safety/ITT population, and have at least one post-baseline disease assessment (at least one post-baseline imaging assessment in accordance with RECIST 1.1, or documented clinical progression [every effort should be made to objectively assess radiographic progression]). Patients who have primary reason for discontinuation as death will be included in the mITT population."

- The PP population was updated from "All patients in the mITT Population who have no major protocol violations during the study, receive at least one cycle (28 doses) of ARQ 087, have at least one post-baseline efficacy measurement, and test positive for FGFR2 gene fusion by the central laboratory" to "will include all patients in the mITT population who have no important protocol (IPD) deviations which are determined to potentially impact efficacy analyses during the study. Protocol deviations will be identified prior to final analysis." This was to reflect the Labcorp standard to use "important" rather than "major", and furthermore to allow flexibility in defining those IPDs that may affect the efficacy endpoints.

- DCR has been added as an exploratory efficacy endpoint, and is defined as the achievement of confirmed CR or PR or SD (or Non-CR/Non-PD for patients with non-measurable disease).

- The minimum duration of SD in the protocol had referenced both 8 weeks (+/-2 days) on page 101 and 8 weeks (+/- 3 days) on page 105 of the protocol. Following client discussion, it was decided that the minimum duration of SD should instead be at least 6 weeks i.e. 42 days as per RECIST guidelines. This minimum duration will also apply to Non-CR/Non-PD.

- Although RECIST does not expect the best overall response to include non-CR/non-PD (NN) for non-measurable disease, in the event that this happens it will be treated as its only entity and in terms of hierarchy it will fall between SD and PD.

- Exploratory endpoints have been added to take into consideration of investigator assessments for ORR, PFS and DCR.

5 REFERENCES

Brookmeyer, R. and Crowley, J. (1982), "A Confidence Interval for the Median Survival Time," Biometrics, 38, 29 - 41.

(Eisenhauer 2009) Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009; 45(2):228-247.

(Kaplan & Meier 1958) Kaplan EL, Meier P. 1958. Nonparametric estimation of incomplete observations. J Am Stat Assoc 53:457-481.

Lamarca A, Hubner RA, Ryder WD, Valle JW. Second-line chemotherapy in advanced biliary cancer: a systematic review Ann Oncol. 2014;25:2328-38

Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika. 1983;70:659-63.

Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are promising: a practical guide with examples. Stat Med. 2011;30(28):3267-84.

(Musoro 2018) Musoro ZJ, Hamel J-F, Ediebah DE, et al. EORTC Quality of Life Group. 2018b. Establishing anchor-based minimally important differences (MID) with the EORTC quality of-life measures: a meta-analysis protocol. BMJ Open;8:e019117.

SAS/Stat User's Guide Version 9.4 (2008). SAS Institute, Cary NC.