EXELIXIS™

# X184-309: Statistical Analysis Plan

A Phase 3, Randomized, Double-blind, Controlled Study of Cabozantinib (XL184) vs Placebo in Subjects with Hepatocellular Carcinoma Who Have Received Prior Sorafenib

**Version 2.0**

Date: 08 Jun 2016

Prepared by: Milan Mangeshkar 08 June 2016
Milan Mangeshkar — Date
Senior Director, Biostatistics, Exelixis

Approved by: 09 JUN 2016
Colin Hessel — Date
Vice President, Development Operations, Exelixis

Approved by: 9 June 2016
Anne Borgman, MD — Date
Vice President, Clinical Development, Exelixis

Approved by: 09 Jun 2016
Gisela Schwab, MD — Date
President, Product Dev and Medical Affairs & CMO, Exelixis

Approved by: 09 June 2016
Lisa Sauer — Date
Vice President, Regulatory Affairs & Quality Assurance, Exelixis

**TABLE OF CONTENTS**

## LIST OF APPENDICES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| AE | Adverse event |
| AFP | Alpha fetoprotein |
| ALP | Alkaline phosphatase |
| ALT | Alanine aminotransferase |
| AST | Aspartate aminotransferase |
| ATA | Adequate tumor assessment |
| ATC | Anatomical Therapeutic Chemical |
| BLQ | Below limit of quantitation |
| BMI | Body mass index |
| BSAP | Bone-specific alkaline phosphatase |
| BUN | Blood urea nitrogen |
| CI | Confidence interval |
| CMH | Cochran-Mantel-Haenszel |
| CRF | Case report form |
| CT | Computerized tomography |
| CTC | Circulating tumor cell |
| CTCAE | Common terminology criteria for adverse events |
| CTMS | Clinical trial management system |
| CTx | C-terminal cross-linked telopeptides of type I collagen |
| DBP | Diastolic blood pressure |
| EBRT | External beam radiation therapy |
| ECOG PS | Eastern Cooperative Oncology Group Performance Status |
| ER | Emergency room visit |
| FDA | Food and Drug Administration |
| GGT | Gamma-glutamyltransferase |
| HBCAB | Hepatitis B core antibody |
| HBsAg | Hepatitis B surface antigen |
| HBV | Hepatitis B virus |
| HCC | Hepatocellular carcinoma |
| HCRU | Health care resource utilization |
| HCV | Hepatitis C virus |

| | |
|---|---|
| HGB | Hemoglobin |
| HbA1c | Glycated hemoglobin |
| HR | Hazard ratio |
| HRQOL | Health Related Quality of Life |
| ICH | International Conference on Harmonization |
| ICU | Intensive care unit |
| IDMC | Independent Data Monitoring Committee |
| INR | Prothrombin International normalized ratio |
| ITT | Intent-To-Treat |
| IVRS | Interactive Voice Response System |
| IWRS | Interactive Web Response System |
| LDH | Lactate dehydrogenase |
| LLN | Lower limit of normal |
| LLQ | Lower limit of quantitation |
| MRI | Magnetic resonance imaging |
| MedDRA | Medical Dictionary for Regulatory Activities |
| NPACT | Non-protocol anti-cancer therapy |
| NTx | N-terminal cross-linked telopeptides of type I collagen |
| OS | Overall survival |
| ORR | Objective response rate |
| PD | Progressive Disease |
| PFS | Progression-free survival |
| PK | Pharmacokinetic |
| PP | Per-protocol |
| qd | Once daily |
| QOL | Quality of Life |
| SAE | Serious adverse event |
| SBP | Systolic blood pressure |
| SAP | Statistical analysis plan |
| SD | Stable Disease |
| TEAE | Treatment emergent-adverse event |
| TSH | Thyroid-stimulating hormone |

| | |
|---|---|
| UE | Unable to evaluate |
| ULN | Upper limit of normal |
| ULQ | Upper limit of quantitation |
| UPCR | Urine protein/creatinine ratio |
| VAS | Visual analogue scale |
| WHO-DD | World Health Organization drug dictionary |

# 1 ADMINISTRATIVE STRUCTURE AND VERSION HISTORY

This study is being conducted under the sponsorship of Exelixis, Inc. Statistical programming and analyses are being conducted under contract by PPD in conjunction with Exelixis, Inc.

This version of the Statistical Analysis Plan (SAP) is based on amendment 1.0 of the protocol dated 23, April 2014.

**Table 1: Protocol Version History**

| Date | Version | Primary Reason(s) for Amendment |
|---|---|---|
| 12 March 2013 | Original Protocol | Not Applicable |
| 23 April 2014 | Amendment 1.0 | Provides additional clarification to 3 inclusion criteria and 4 exclusion criteria.<br><br>Additional Child-Pugh testing time-points have been added.<br><br>Specifies that all subjects will have hepatitis virus testing via central laboratory prior to enrollment but those results will not be required for randomization.<br><br>Introduces the Maintenance Phase, which subjects will enter when sufficient data have been collected to evaluate all study endpoints. |

**Table 2: SAP Version History**

| Date | Version | Primary Reason(s) for Amendment |
|---|---|---|
| 2015NOV02 | Original | |
| 2016JUN08 | Version 2.0 | All efficacy analyses will be w.r.t IxRS stratification factors instead of stratification factors based on CRF<br><br>Per protocol population deleted as analyses by ITT population is deemed more robust<br><br>Baseline for biomarkers defined w.r.t first dose date as |

| Date | Version | Primary Reason(s) for Amendment |
|---|---|---|
| | | samples were collected prior to first dose date and not prior to randomization<br><br>Surgical history summarization deleted as all fields collected on the CRF are free text intended to support subject narratives as needed<br><br>Progression for a subject updated to accept new lesions identified on bone scan as evidence of progression. Adequate tumor assessments will include unscheduled tumor assessments<br><br>Refined PFS sensitivity analyses (PFS2 and PFS3) algorithm to exclude treatment discontinuation due to adverse event<br><br>Repeated measures analysis added for EQ-VAS and EQ-Index parameters. Per patient reported outcome experts, revised the definition of potentially clinically meaningful effect size from 0.3 – 0.5 to ≥ 0.3<br><br>Health care resource utilization section was added for summarizing hospitalizations, emergency room visits, intensive care unit visits<br><br>Subgroup categories refined to include only relevant prognostic factors<br><br>Safety observation period refined to better reflect a subject's time on study treatment<br><br>Exposure summary will be in months instead of weeks<br><br>Blood pressure summary revised to consider treatment |

| Date | Version | Primary Reason(s) for Amendment |
|---|---|---|
| | | emergent adverse finding |

# 2 STUDY DESCRIPTION

## 2.1 Study Design

This is a Phase 3 multicenter, randomized, double-blinded, controlled trial of cabozantinib vs. placebo, both with best supportive care, in subjects with advanced hepatocellular carcinoma (HCC) who were previously treated with sorafenib. The primary efficacy endpoint for the study is overall survival (OS). Approximately 760 subjects will be randomized in a 2:1 ratio to receive either cabozantinib or placebo.

Each subject's course of treatment will consist of the following periods:

**Pre-treatment Period**: Potential subjects will be screened to determine if they meet the required eligibility criteria. Qualifying screening assessments must be performed within 28 days before randomization unless otherwise specified.

**Treatment Period**: Subjects who meet all study eligibility criteria, in addition to receiving best supportive care, will be randomly assigned in a 2:1 ratio to the cabozantinib and placebo arms respectively. Crossover between treatment arms will not be allowed.

Subjects will receive blinded study treatment and best supportive care as long as they continue to experience clinical benefit in the opinion of the investigator or until there is unacceptable toxicity or the need for subsequent systemic anti-cancer treatment or liver-directed local anti-cancer therapy. Study treatment may even continue after radiographic progression as long as the investigator believes that the subject is still receiving clinical benefit and that the potential benefit of continuing study treatment outweighs potential risk.

When sufficient data have been collected to evaluate all study endpoints and upon site notification from the Sponsor, subjects remaining on study treatment will enter the study Maintenance Phase. In this phase subjects will continue to receive study treatment until a criterion for protocol-defined discontinuation has been met. Subjects will undergo safety assessments and tumor assessments per standard of care.

**Post-Treatment Period**: The final safety assessment will occur at the post-treatment follow-up visit 30 (+14) days after the date of the decision to discontinue treatment unless a Grade 3/4 AE or an SAE is determined to be ongoing.

Radiographic tumor assessments and EQ-5D-5L quality of life (QOL) assessments will continue per the protocol-defined schedule, regardless of whether study treatment is given, reduced, interrupted, or discontinued until the later of 8 weeks after radiographic progression per RECIST 1.1 as determined by the investigator or the date of the decision to permanently discontinue study treatment.

Subjects will additionally be contacted approximately every 8 weeks after the post-treatment follow-up visit to assess survival status and to document receipt of subsequent anti-cancer therapy. Every effort will be made to collect these protocol-specific evaluations unless consent to participate in the study is withdrawn.

The study design, schedule of the visits, assessments and conduct are described in the study protocol.

## 2.2 Study Treatment

Eligible subjects will be randomized in a 2:1 ratio to the following treatment arms:

- Cabozantinib arm: Oral cabozantinib (60 mg) qd
- Placebo arm: Oral cabozantinib-matched placebo qd

In addition, all subjects will also receive best supportive care.

## 2.3 Study Objectives and Endpoints

The primary objective of this study is to evaluate the effect of cabozantinib compared with placebo on overall survival in subjects with advanced HCC previously treated with sorafenib.

### 2.3.1 Primary Efficacy Endpoint

The primary efficacy endpoint is overall survival (OS) (see section 7.1.1for definition).

### 2.3.2 Secondary Efficacy Endpoint

The secondary efficacy endpoints are:

- Objective response rate (ORR) per RECIST 1.1 (see section 7.1.1)
- Progression-free survival (PFS) per RECIST 1.1 (see section 7.1.1)

### 2.3.3 Additional Endpoints

The following endpoints are discussed in their respective sections:

- Safety and tolerability (see section 8)
- Relationship of baseline and changes in biomarkers with treatment and/or clinical outcome (see section 7.5)
- Health-related quality of life (HRQOL) as assessed by the EuroQol Health questionnaire instrument (EQ-5D-5L) (see section 7.5.1)
- Pharmacokinetics (PK) (see section 7.5.7)

## 2.4 Power and Sample Size Justification

For the primary OS endpoint, up to three event-driven analyses are planned at 50%, 75%, and 100% information fraction (311, 466, and 621 deaths, respectively). A sample size of 760 subjects with a total of 621 events (and two interim analyses) provides the study with 90% power for a 2-sided log-rank test at 5% level of significance to detect a 31.6% increase in OS (HR = 0.76). Assuming a median OS of 8.2 months in the placebo arm (based upon the BRISK trial; Llovet 2012) and exponential distribution, this corresponds to median OS of 10.8 months in the cabozantinib arm.

In the current design, the minimum observed effect that would result in statistical significance for OS at the two interim and final analyses are 42.1% improvement (HR = 0.70, from 8.2 to 11.7 months), 25.7% improvement (HR = 0.80, from 8.2 to 10.3 months) and 18.4% improvement (HR = 0.84, from 8.2 to 9.7 months), respectively.

With an average accrual rate of 31.5 subjects per month and using a 2:1 treatment allocation ratio, an approximate total of 760 subjects (507 subjects in cabozantinib arm and 253 subjects in placebo arm) are required to observe the minimum number of events within the planned study duration (25 months accrual and approximately 38 months to observe the required events for OS). Currently since the average accrual rate is less than 31.5 subjects per month this will also prolong the study duration and time required to observe the required events for OS.

Power and sample size estimates were calculated using EAST v5 by Cytel Software.

## 2.5 Randomization and Blinding

This is a randomized, double-blinded, controlled trial of cabozantinib versus placebo. cabozantinib-matched placebo will be packaged and color-, size-, and shape-matched to be indistinguishable from cabozantinib.

Study treatment assignment will be unknown to the subjects, investigators, study centers, the Sponsor, and any contract research organization affiliated with the study other than those authorized to access treatment assignment for regulatory safety reporting and submission processes (see Section 8.2.2 of Protocol), IVRS or IWRS system administration and drug supply management.

When an individual subject has been deemed eligible at the study site, the site representative will use the designated interactive voice response system/interactive web response system (IVRS/IWRS) to enroll the subject into the study. Eligible subjects will be randomly assigned in a 2:1 ratio to the cabozantinib and placebo arms.

Randomization will be stratified by the following factors:

- etiology of disease (HBV [with or without HCV], HCV [without HBV], or Other),
- geographic region (Asia [includes Japan, Hong Kong, South Korea, Singapore, Malaysia, Philippines, Taiwan, Thailand, Vietnam], Other Regions)
- presence of extrahepatic spread of disease and/or macrovascular invasion (Yes, No).

Note that IVRS/IWRS will be referred as IxRS in this document and HBV refers to hepatitis B virus and HCV refers to hepatitis C virus.

# 3 ANALYSIS POPULATIONS

## 3.1 Intent to Treat

The Intent-To-Treat (ITT) population is defined as all randomized subjects regardless of whether any study treatment or the correct study treatment was received. This population will be used for efficacy analyses.

### 3.2 Safety Population

The Safety population will include all randomized subjects who receive any amount of study treatment (either cabozantinib or cabozantinib-matched placebo). Analyses based on the Safety population will be performed according to the actual treatment received. Subjects randomized to the placebo arm who receive any amount of cabozantinib in error will be summarized in the cabozantinib group.

### 3.3 Per Protocol Population

A Per Protocol population is not defined or planned for the study.

## 4 GENERAL CONVENTIONS

The statistical principles applied in the design and planned analyses of this study are consistent with International Conference on Harmonization (ICH) E9 guidelines (ICH 1998).

Continuous data will be summarized using descriptive statistics (number of observations, mean, standard deviation, median, $25^{th}$ and $75^{th}$ percentiles, minimum, and maximum). Frequencies and percentages will be used for summarizing categorical (discrete) data.

Confidence intervals, when presented, will generally be constructed at the 95% level. For binomial variables, the normal approximation methods will be employed unless otherwise specified.

A month is operationally defined to be 30.4375 days. Six months is operationally defined to be 183 days.

All summaries will be presented by treatment arm unless otherwise specified.

### 4.1 Definition of Baseline

In general, for efficacy endpoints the last observed measurement prior to randomization will be considered the baseline measurement. Exception to this rule are efficacy markers such as pharmacogenetics blood samples, biomarker samples, bone marker samples and blood samples for potential circulating tumor cell (CTC) analyses. These samples per schedule of assessment were collected prior to the first dose date hence the baseline measurements will be with respect to the first dose date. For subjects who did not take any study treatment, any

biomarker sample available prior to randomization will be considered as baseline observation.

For safety endpoints the last observation before first dose of study treatment will be considered the baseline measurement unless otherwise specified. For assessments on the day of first dose where time is not captured, a nominal pre-dose indicator, if available, will serve as sufficient evidence that the assessment occurred prior to first dose.

Assessments on the day of first dose where neither time nor a nominal pre-dose indicator are captured will be considered prior to first dose if such procedures are required by the protocol to be conducted before first dose.

## 4.2 Definition of Study Day

For the purpose of efficacy data summaries, study day is defined with respect to the randomization date. For visits (or events) that occur on or after randomization, study day is defined as (date of visit [event] – date of randomization + 1). For visits (or events) that occur prior to randomization, study day is defined as (date of visit [event] – date of randomization). There is no Study Day 0.

For the purpose of safety data summary, Dose Day 1 is defined as the date of first dose of study treatment (referred to in the protocol as Week 1 Day 1). For visits (or events) that occur on or after first dose, dose day is defined as (date of visit [event] – date of first dose of study treatment + 1). For visits (or events) that occur prior to first dose, dose day is defined as (date of visit [event] – date of first dose of study treatment). There is no Dose Day 0.

For listings (such as for adverse events [AEs]) that include the derivation of "days since last dose," this is defined as (event date – date of last dose). Events that occur on the same day as the last dose of study drug will therefore be described as occurring zero days from the last dose of study drug.

## 4.3 Visit Window Calculation

Analyses will be according to actual visit dates and times. The planned analyses do not require the calculation of visit windows. However, for analyses that require a particular visit as planned, measurements included in that visit must have occurred during the acceptable window defined for the visit in the protocol, unless otherwise specified in this plan.

## 4.4 Missing and Partial Data

In general, other than for partial dates, missing data will not be imputed and will be treated as missing. The algorithms for imputation of partial dates vary depending upon the parameter. These are presented in Appendix A.

## 4.5 Safety Observation Period

The safety observation period is defined as time from first dose date of study treatment to the earlier of the date of the decision to permanently discontinue study treatment+ 30 days or date subject withdrew consent or date of death or data cut-off date.

Generally only the safety data (including adverse events, laboratory results, vital signs, ECG, ECOG PS, concomitant medications and etc.) reported during the safety observation period will be analyzed and summarized, unless otherwise specified in this plan.

## 4.6 Definition of Prior, Concomitant, and Subsequent Therapy

For the purpose of inclusion in summary tables, incomplete medication or radiation start and stop dates will be imputed as detailed in Appendix A. Based on imputed start and stop dates:

- Prior medications/radiation therapies are defined as medications with a start date occurring before the date of first dose of study treatment.
- Concomitant medications/radiation therapies are defined as medications that stop or continue on or after date of first dose day through the end of safety observation period.
- Subsequent medications/radiation therapies are defined as those that stop or continue on or after the date of randomization.
- Medications/radiation therapies may be summarized as prior, concomitant and/or subsequent.

## 4.7 Software

All analyses will be conducted using SAS Version 9.1 or higher.

### 4.8 Changes to Planned Analyses

Substantive changes to the analyses described in the protocol or in approved versions of this plan will be fully documented in a revised version of the plan approved by the Sponsor prior to unblinding the study to conduct the analyses.

Clarifications, minor corrections, and operational considerations necessary to accurately conduct the analyses that do not materially change the nature of the analysis will be documented in an addendum to this plan that will be also be approved by the Sponsor prior to unblinding the study to conduct the analyses.

## 5 STUDY POPULATION SUMMARIES

### 5.1 Enrollment

Subjects are defined to be enrolled at randomization. Enrollment will be summarized by region, country, site and protocol amendments.

### 5.2 Disposition

Subject disposition will be summarized categorically and will include the number and percentage of subjects in the ITT and Safety populations.

The reasons for study treatment discontinuation and study follow-up discontinuation will also be summarized categorically.

### 5.3 Demographic and Baseline Characteristics

Summaries of demographics, stratification factors and baseline characteristics will be presented for subjects in the ITT and Safety populations.

[A] The demographic characteristics include:

- Age (continuous)
- Age category 1
  - < 65 years
  - ≥ 65 years
- Age category 2
  - < 75 years
  - >=75 years

- Age category 3
  - <65 years
  - 65 to <75 years
  - 75 to <85 years
  - ≥85 years
- Sex
  - Male
  - Female
  - Not reported
- Ethnicity
  - Hispanic or Latino
  - Not Hispanic or Latino
  - Not Reported
- Race
  - American Indian or Alaska Native
  - Asian
  - Black or African American
  - Native Hawaiian or Other Pacific Islander
  - White
  - Not Reported
  - Other
- Geographic Region
  - Australia/New Zealand
  - Asia (excluding Japan)
  - Europe
  - North America (Canada/USA)
- ~~o~~ Japanese decent (world-wide)
  - Japanese descent
  - Non-Japanese descent
  - Both Japanese and Non-Japanese descent
  - Not Reported

Note for this study birth date is not collected but age in years is collected at informed consent.

[B] Categorical summaries of the following stratification factors will be presented as recorded (a) in the IxRS during randomization (b) on the CRF (c) cross tabulation of all 3 stratification factors per IxRS (d) cross tabulation of all 3 stratification factors per CRF (e) cross-tabulation of geographic region and etiology of disease per CRF:

- etiology of disease
  - HBV [with or without HCV]
  - HCV [without HBV]
  - Other
- geographic region (Asia, Other Regions)
- presence of extrahepatic spread of disease and/or macrovascular invasion (Yes, No)

[C] Baseline characteristics include:

- Height in inches– descriptive statistics
- Weight in kg – descriptive statistics
- Body mass index (BMI) in kg/meter$^2$, calculated as (weight in kg*1000)/(Height in cm)$^2$ –descriptive summary:
- ECOG PS: 0, 1, Missing
- Smoking history
  - Current
  - Former
  - Never
- Alcohol use – Categorical summary for subjects classified as current user, former user or never will be presented

[D] Descriptive statistics and or categorical summaries for the following baseline laboratory characteristics :

- Alpha fetoprotein [AFP] ( $< 400$ ng/ml, $\geq 400$ ng/ml)
- Prothrombin International normalized ratio [INR] ($\leq 2.3$, $>2.3$)
- Albumin ($< 35$ g/L, $\geq 35$ g/L)
- Total bilirubin in umol/L ($< 22.23$, $\geq 22.23 – 29.07$, $\geq 29.07$)
- Neutrophil/Lymphocyte ratio ($< 3$, $\geq 3$)

- Albumin and total bilirubin (ALBI) grade derived from calculated ALBI score = [log(Bilirubin in umol/L)*0.66 – (Albumin in g/L)*0.085]
    - Grade 1: ALBI score ≤ -2.60
    - Grade 2: -2.60 < ALBI score ≤ -1.39
    - Grade 3: ALBI score > -1.39

For laboratory parameters summarized as baseline characteristics, the most-recent non-missing central or local sample available before the date and time of randomization will be employed. This differs from definition of baseline laboratory values used in safety summaries.

## 5.4 Medical History

General medical history data will be coded per MedDRA.

## 5.5 Cancer History and Current Disease Status

Cancer history and current disease characteristics data collected on the cancer history CRF will be summarized categorically or with descriptive statistics as appropriate. The following summaries are planned:

- Diagnosis of carcinoma of HCC by histology or cytology (Yes, No)
- Current etiology:
    - Hepatitis B virus (without HCV)
    - Hepatitis C virus (without HBV)
    - Hepatitis B and C virus
    - Hepatitis B virus (regardless of HCV)
    - Hepatitis C virus(regardless of HBV)
    - Alcoholism
    - Nonalcoholic Steatohepatitis (NASH)
    - Other
- Hepatitis B virus surface antigen (HBsAg), Hepatitis B virus core antibody (HBCAB) and HCV laboratory results summary
    - HBsAg(+) , HBCAB(+)
    - HBsAg(+) , HBCAB(-)
    - HBsAg(-) , HBCAB(+)
    - HBsAg(-) , HBCAB(-)

    - HCV(+)
    - HCV(-)
- Child-Pugh Grade
    - A (score 5 - 6)
    - B (score 7 – 9)
    - C (score 10 – 15)
- Hepatic Encephalopathy
    - None
    - Grade I-II
    - Grade III-IV
    - Missing or unknown
- Ascites
    - Absent
    - Slight
    - Moderate
    - Missing or unknown
- Time in years to randomization since diagnosis of HCC as identified by histology or cytology (Note: Incomplete diagnosis dates will be imputed as detailed in Appendix A)
- Currently has locally advanced disease (Yes, No, Unknown)
- Currently has metastatic disease (Yes, No)
- Extent of disease at baseline per target/non-target sites identified by the investigator on the tumor assessment CRFs (liver, lymph node, brain, bone, other, peritoneum , visceral sites other than liver , etc.)
- Number of target/non-target sites at baseline identified on tumor assessment CRFs per investigator (1, 2, $\geq$ 3)
- Has bone metastasis at baseline per history of bone lesion CRF or as identified by target and non-target lesion CRFs (Yes, No)
- Has measurable disease at baseline identified by the investigator on tumor assessment CRFs (Yes, No)
- Current Extent of HCC Disease :
    - Portal Vein Invasion (Yes, No, Unknown)
    - Bile Duct Invasion (Yes, No, Unknown)
    - Macrovascular Invasion (Yes, No, Unknown)

- Extrahepatic Spread (Yes, No, Unknown)
- Other (Yes, No)

- MET immunohistochemistry status (High, Low/Negative, Unknown). The status of high and low/negative will be based on cutoff of ≥50% of tumor tissue stained with an intensity of 2+ or 3+). The cut-off is based on historical NSCLC and HCC data, may be adjusted if warranted based on results in initial XL184-309 data transfers
- VEGF-A amplification in circulating tumor cells (Yes, No, Unknown)

# 6 TREATMENTS AND MEDICATIONS

## 6.1 Prior Anti-Cancer and Radiation Therapy

Prior anti-cancer therapies will be coded per World Health Organization drug dictionary (WHO-DD).

The following will be summarized categorically or with descriptive statistics as appropriate for all subjects in the ITT and Safety population:

- Received prior sorafenib for HCC (Yes, No)
- Descriptive summary for duration of treatment in months on prior sorafenib for HCC
- Categorical summaries for duration of treatment in months on prior Sorafenib for HCC as follows:
  - < 1 month, ≥ 1 month
  - < median, ≥ median
- Number and percent of subjects who progressed on the most recent prior systemic non-radiation anti-cancer therapy for HCC (Yes, No)
- Number and percent of subjects who got prior sorafenib as the most recent prior anti-cancer therapy for HCC and progressed (Yes, No)
- Cross tabulation of type of non-radiation therapy (Local, Systemic and Unknown) with indication (for HCC [Local-Liver directed, Local-Other locations, Adjuvant, Advanced, Other], Other than HCC)
- Number of prior systemic non-radiation anti-cancer regimens for advanced HCC per subject (1, 2, ≥3) and descriptive statistics

- Number of prior systemic non-radiation anti-cancer agents for advanced HCC per subject (1, 2, ≥3) and descriptive statistics
- The time from the end of most-recent non-radiation prior systemic anti-cancer treatment for HCC to randomization will be summarized descriptively Number of prior radiation therapies for HCC per subject (1, 2, ≥3) and descriptive statistics
- Subject incidence of radiation therapy by indication from history of radiation therapy (Disease under study and Other)
- Subject incidence of radiation therapy type (External beam radiation therapy [EBRT], Internal radiation therapy [brachytherapy], Radioisotope therapy, Radioembolization, Radiofrequency ablation and Other) from history of radiation therapy received for HCC or Other indications
- Subject incidence of site (Bone, Soft-tissue, Systemic, Unknown) of radiation from history of radiation therapy received for HCC or Other indications

All prior non-radiation anti-cancer agents will be summarized categorically by Anatomical Therapeutic Chemical (ATC) Class Text and WHO-DD base substance preferred name by treatment arm and type of therapy (Local-liver directed, Local-non-liver directed, Systemic, Unknown) for all subjects in the ITT population.

### 6.2 Prior and Concomitant Medications

Medications recorded on the CRFs will be coded using the WHO-DD. Prior and concomitant medications, other than prior and subsequent anticancer therapies will be summarized by treatment group in the Safety population by ATC and WHO-DD base substance preferred name. In addition, prior medication will also be summarized in the ITT population by ATC and WHO-DD base substance preferred name. Anticancer therapies are addressed in sections 6.6 and 6.7 of this plan.

### 6.3 Study Treatment Exposure

Study treatment exposure will be summarized with descriptive statistics in the Safety population.

The following will be derived for each subject and will be summarized:

- Duration (in months) of exposure per subject, calculated as (date of decision to discontinue study treatment – date of first dose + 1) /30.4375

- Average daily dose per subject (mg/day) of cabozantinib (or cabozantinib-matched placebo), calculated as (total dose received / duration of exposure)
- Percent dose intensity for cabozantinib (or cabozantinib-matched placebo) calculated as 100*(average daily dose mg/day) / (60 mg/day)
- Duration of treatment in months defined as (date of decision to discontinue study treatment – date of first dose – total duration of dose holds + 1) /30.4375

## 6.4 Study Treatment Modifications

Treatment modifications (holds and reductions) for cabozantinib (or cabozantinib-matched placebo) will be summarized in the Safety population. Only modifications due to AE will be summarized.

A. The following summaries will be presented for the cabozantinib (or matched placebo) component of study treatment:

   i. For dose reductions due to AE

      Categorical summaries for:

      - Subjects with any dose reduction
      - Dose levels received by a subject
      - Lowest non-zero dose level received
      - Last non-zero dose level received
      - Last dose level received (including dose holds)

      Descriptive statistics for:

      - Duration of treatment in months for each dose level (60 mg, 40 mg, 20 mg, 0 mg)
      - Time to first dose level reduction (first receipt of 40mg) (days)
      - Time to second dose level reduction (first receipt of 20mg) (days)

   ii. Summaries for dose holds due to AE:
      - Descriptive statistics for number of dose holds due to an AE
      - Descriptive statistics for duration of dose holds per dose hold and per subject due to an AE, calculated as (stop date of hold – start date of hold + 1)
      - Categorical summary for subjects with duration of holds due to an AE that can be classified as any number of days, $\geq 7$ days, $\geq 14$ days, and $\geq 21$ days

- Descriptive statistics for time to first dose hold, time to first dose hold that was ≥ 7 days, ≥ 14 days, and ≥ 21 days. The time to dose hold is calculated as (start date of the hold – first dose date + 1)
- Descriptive statistics for time to second dose hold, time to second dose hold that was ≥ 7 days, ≥ 14 days, and ≥ 21 days

iii. Summaries for dose modifications (defined as a reduction or hold) due to AE:
- Frequency counts and percentages for subjects with any dose modifications
- Descriptive statistics for number of dose modifications (0-3)
- Descriptive statistics for time to the first dose modification
- Descriptive statistics for time to the second dose modification

## 6.5 Study Treatment Non-Compliance and Dosing Errors

Treatment non-compliance and dosing errors for reasons other than AE will be summarized in the Safety population. Frequency counts and percentages will be presented by treatment groups for:

- Subjects with dose hold (0 mg) due to non-compliance
- Subjects who received dose > maximum allowed dose level at any time
- Subjects who received wrong dose (≤ maximum allowed dose level) at any time due to non-compliance
- Subjects who received wrong dose (≤ maximum allowed dose level) at any time due to site/logistic error or other reason

## 6.6 Non-Protocol Anti-cancer Therapy

For the purpose of supporting safety evaluation:

Non-radiation concomitant (see definition in section 4.6.) NPACT will be summarized by ATC text and WHO Drug base substance preferred name in the Safety population.

For the purpose of supporting efficacy evaluation:

All subsequent (see definition in section 4.6.) NPACT, including radiation therapy, will be summarized by treatment group in the ITT population as follows:

- Based on the non-radiation therapy received subjects will be categorized into one or more of the following categories: systemic, local, or unknown and all NPACTs falling under these categories will be summarized by ATC text and WHO Drug based substance preferred name
- Time to first systemic NPACT will be summarized by descriptive statistics
- Frequency counts and percentages will be presented for radiation therapy indication, type and site

### 6.7 Post-randomization Surgery/Procedure

Post-randomization surgery/procedures that impacted the tumor lesion(s) (Yes, No, Unknown) will be summarized by treatment group for subjects in the ITT and Safety populations.

### 6.8 Concomitant Transfusions

Concomitant transfusions will be summarized by transfusion type and treatment group for subjects in the safety population.

## 7 EFFICACY ANALYSES

Primary efficacy analyses will be performed on all subjects in the ITT population.

### 7.1 Primary Efficacy Endpoint (OS)

#### 7.1.1 Definition

The primary efficacy endpoint is overall survival. Duration of OS is defined as the time from randomization to death due to any cause. For subjects who are alive at the time of data cutoff but are permanently lost to follow-up, duration of OS will be right censored at the date the subject was last known to be alive. Those who withdraw consent from survival follow-up and are alive will be right censored at the date the subject withdrew consent from survival follow-up. Subjects alive on or after the data cutoff or those who died after the data cutoff will be right censored at the date of data cutoff.

OS (months) = (earliest date of death or censoring – date of randomization + 1)/30.4375

### 7.1.2 Hypothesis

The hypotheses to be evaluated in the analysis of the primary efficacy endpoint are:

$H_0$: $S(t)_{cabozantinb} = S(t)_{placebo}$

$H_A$: $S(t)_{cabozantinb} \neq S(t)_{placebo}$

where $S(t)_{Cabozantinib}$ and $S(t)_{Placebo}$ are the survivor functions for the cabozantinib and placebo arms, respectively.

### 7.1.3 Primary Analysis

The primary analysis of OS will include all subjects in the ITT population.

The hypothesis testing between the two treatment arms will be performed using the stratified log-rank test with a 2-sided $\alpha$=0.05 level of significance. The stratification factors are as described in Section 2.5 and the values used for analysis will be those recorded in the (IxRS).

The median duration of OS and the associated 95% confidence interval (CI) for each treatment arm will be estimated using the Kaplan-Meier method. The stratified hazard ratio (HR) and its 95% CI will be estimated using a Cox proportional-hazard model with treatment group as the independent variable and stratified by the same randomization stratification factors as were used for the log-rank test.

The analysis of OS is event-based. Up to three analyses of OS are planned: two interim analyses and a final analysis that will occur when 311, 466 and 621 deaths (i.e. 50%, 75% and 100% deaths) have been observed respectively. Inflation of Type I error associated with interim analyses will be controlled using a Lan-DeMets O'Brien-Fleming alpha-spending function. The critical p-values (and observed HR) for rejecting the null hypothesis will be 0.0031 (HR$\leq$ 0.70), 0.0183 (HR$\leq$ .80) and 0.044 (HR$\leq$ .84) at the time when 311, 466 and 621 deaths (50%, 75% and 100% information) are observed respectively. The actual critical values will depend upon the true number of events observed at each analysis. The interim analyses of OS are also described in section 7.4.

At an analysis time point, if the p-value for the stratified log-rank test is less than the critical value for rejecting the null hypothesis and the HR is < 1, the null hypothesis will be rejected

and it will be inferred that OS is superior in the cabozantinib arm compared to the placebo arm.

### 7.1.4 Exploratory Analyses

Overall survival analyses as described in section 7.1.3 will be conducted by censoring for subjects who receive a systemic NPACT or a local liver directed therapy after randomization.

## 7.2 Secondary Efficacy Endpoints

The secondary efficacy endpoints for this study are duration of PFS and ORR. Formal hypothesis tests are planned for the secondary efficacy endpoints.

### 7.2.1 Progression-Free Survival (PFS)

#### 7.2.1.1 Definition

Duration of PFS is defined as the time from randomization to the earlier of the date of radiographic progression or date of death due to any cause.

PFS (months) =(earliest date of progression, death, censoring – date of randomization + 1)/30.4375

The primary analysis of PFS will include all subjects in the ITT population and will include radiographic progression events as determined by the investigator per RECIST 1.1 and deaths.

The recorded date of radiographic progression is the date of the tumor assessment visit at which progression is declared. If multiple scan dates are associated with a tumor assessment visit, the earliest assessment date within the set will be chosen as the progression date.

Only adequate tumor assessments (ATAs) will be considered in the determination of radiographic progression and censoring dates. For the purpose of this study, ATA is defined as one that results in a time point assignment of: response (complete or partial), stable disease/(non-CR, non-PD), or progression. For PFS, ATA is based on soft tissue evaluation by CT/MRI and/or new lesions identified by bone scan.

General censoring rules for the primary analysis of PFS are described below:

- Subjects who receive systemic or liver directed local NPACT or non-protocol radiation therapy (other than to bone) or surgery to resect tumor lesions before experiencing an

event will be right censored at the date of the last ATA on or prior to the date of initiation of subsequent therapy/surgery. If there is no such tumor assessment after randomization, the subject will be right censored on the date of randomization.

- Subjects who have not experienced an event (and are not otherwise censored) at the time of data cutoff will be right censored on the date of their last tumor assessment after randomization that is on or prior to the data cutoff. If there is no such tumor assessment after randomization, the subject will be right censored on the date of randomization.

- Subjects who miss two or more ATAs (operationally defined as 126 days without an ATA) followed by an event (progression or death) will be right censored on the date of their most-recent ATA prior to the missing assessments. If there is no such tumor assessment after randomization, the subject will be right censored on the date of randomization.

#### 7.2.1.2 Hypothesis

The hypotheses to be evaluated in the analysis of the PFS are:

$H_0$: $S(t)_{cabozantinb} = S(t)_{placebo}$

$H_A$: $S(t)_{cabozantinb} \neq S(t)_{placebo}$

where $S(t)_{Cabozantinib}$ and $S(t)_{Placebo}$ are the survivor functions for PFS for the cabozantinib and placebo arms, respectively.

#### 7.2.1.3 Primary Analysis

The hypothesis testing of PFS between the two treatment arms will occur only if the result of either an interim analysis or the final analysis of OS achieves statistical significance. The primary analysis of PFS will include all subjects in the ITT population.

The summaries (median and 95% CI, stratified and unstratified log-rank p-values, stratified and unstratified HRs and their 95% CI) and graphs described in section 7.1.3 will be generated for PFS using a 2-sided level of significance, $\alpha$=0.04.

If the p-value for the stratified log-rank test for PFS is less than 0.04 and the HR is < 1, the null hypothesis will be rejected and it will be inferred that PFS is superior in the cabozantinib arm compared with the placebo arm.

#### 7.2.1.4 Sensitivity Analyses

The primary analysis (PFS1) and 2 types of sensitivity analyses (PFS2 and PFS3) are outlined in Table 3. These analyses will include all subjects in the ITT population. Summaries and graphs as described in section 7.1.3 will be presented.

The sensitivity analyses (PFS2 and PFS3) define additional clinical outcomes as events and also evaluate the impact of informative censoring.

For PFS2 analysis, the following are considered as events:

- Radiographic progression per RECIST 1.1
- Death due to any reason
- Treatment discontinuation due to clinical deterioration
- Receipt of systemic non-protocol anti-cancer therapy (NPACT)
- Receipt of local liver-directed NPACT
- Radiation (other than to bone)
- Surgery to resect tumor lesions

For PFS3 analysis, the following are considered as events:

- Radiographic progression per RECIST 1.1
- Death due to any reason
- Treatment discontinuation due to clinical deterioration

**Table 3: Event and censoring rules for primary and sensitivity analyses of PFS**

| Count as censored or event at earliest outcome criterion met | PFS1 | | PFS2 | | PFS3 | |
|---|---|---|---|---|---|---|
| **Purpose** | **Primary** | | **Sensitivity** | | **Sensitivity** | |
| **Situation** | **Outcome** | **Date** | **Outcome** | **Date** | **Outcome** | **Date** |
| No post baseline assessment | Censored | Date of randomization | Censored | Date of randomization | Censored | Date of randomization |
| Radiographic PD | Event | Date of PD | Event | Date of PD | Event | Date of PD |
| Death | Event | Date of death | Event | Date of death | Event | Date of death |
| Subsequent systemic or local liver directed non-protocol anti-cancer therapy (NPACT) | Censored | Date of last ATA on or prior to NPACT | Event | Date of NPACT | Censored | Date of last ATA on or prior to NPACT |
| Radiation (other than to bone) | Censored | Date of last ATA on or prior to Radiation | Event | Date of radiation | Censored | Date of last ATA on or prior to Radiation |
| Surgery to resect tumor lesions | Censored | Date of last ATA on or prior surgery | Event | Date of surgery | Censored | Date of last ATA on or prior surgery |
| Event after more than two missed ATAs (>126 days) | Censored | Date of last ATA prior to the missing visits | Censored | Date of last ATA prior to the missing visits | Censored | Date of last ATA prior to the missing visits |
| Treatment Discontinuation due to Clinical deterioration | NA | NA | Event | Date of determination | Event | Date of determination |
| No Event by last ATA | Censored | Date of last ATA | Censored | Date of last ATA | Censored | Date of last ATA |

ATA:Adequate tumor assessment; PD=Progressive Disease; NA=Not Applicable

### 7.2.2 Objective Response Rate (ORR)

#### 7.2.2.1 Definition

For each subject, best overall response (BOR) is defined as the best tumor assessment category as determined by the investigator per RECIST 1.1 that occurs prior to any of the censoring events defined for the primary analysis of PFS as described in section 7.2.1.1. Tumor assessment categories are ranked as: confirmed complete response (CR), confirmed partial response (PR), stable disease (SD), progressive disease (PD) and not evaluable (NE). To be classified as confirmed CR or confirmed PR, confirmation must have occurred on a subsequent visit that is ≥ 28 days after the response was first observed.

The ORR is defined as the proportion of subjects with a BOR of confirmed CR or confirmed PR.

#### 7.2.2.2 Hypothesis

The hypotheses to be evaluated in the analysis of the ORR are as follows:

$H_0$: $ORR_{Cabozantinib} \leq ORR_{Placebo}$

$H_A$: $ORR_{Cabozantinib} > ORR_{Placebo}$

where $ORR_{Cabozantinib}$ and $ORR_{Placebo}$ are the ORRs for the cabozantinib and placebo arms, respectively.

#### 7.2.2.3 Primary Analysis

Hypothesis testing for ORR will be performed using the Fisher's exact test at the 2-sided $\alpha$=0.01 level of significance. If a sufficient number of responders are observed, analysis using the Cochran-Mantel-Haenszel (CMH) method to adjust for stratification factors per IxRS may also be conducted. The testing of ORR will occur only if the result of either an interim analysis or the final analysis of OS achieves statistical significance. The primary analysis of ORR will include all subjects in the ITT population.

Point estimates of ORR for each treatment arm, the difference in ORR between the two treatment arms, and associated confidence intervals will be provided. The odds ratio and its confidence intervals will also be shown. The 95% CIs for the point estimate will be calculated using exact methods. The 95% CIs for the difference in ORR between the two treatment arms and for the odds ratio will be calculated by asymptotic methods.

If the p-value for the two-sided Fisher's exact test is less than 0.01 and the point estimate for ORR in the cabozantinib arm is higher than that in the placebo arm, the null hypothesis of no difference in ORR will be rejected and it will be inferred that ORR is superior in the cabozantinib arm compared with the placebo arm.

#### 7.2.2.4 Supportive Analyses

To provide a more detailed understanding of anti-tumor activity, the maximum tumor reduction since baseline in target lesions will be derived for those subjects who have baseline and at least one post-baseline measure. The maximum percent tumor reduction from baseline in target lesions for each arm will be displayed graphically using waterfall

plots. For each subject, data from time points after the first date of any of the censoring events defined for the primary analysis of PFS as described in section 7.2.1.1. will be excluded from the waterfall plots.

## 7.3 Control of Type I Error

The multiplicity issue resulting from analysis of one primary endpoint (OS), two secondary efficacy endpoints (PFS and ORR), and planning two interim analyses for testing OS will be addressed by employing a fixed-sequence testing procedure, applying a modified Bonferroni procedure (dividing the alpha between the secondary endpoints), and implementing an alpha-spending function.

Up to three event-driven analyses of OS are planned: two interim analyses and a final analysis (see details in section 7.1.3). Inflation of Type 1 error associated with interim analyses will be controlled using a Lan-DeMets O'Brien-Fleming alpha-spending function.

Interim analysis of PFS and ORR are not planned. The testing of these two secondary endpoints will occur only if the result of either an interim analysis or the final analysis of OS achieves statistical significance. The hypothesis for PFS and ORR will be tested in parallel. PFS will be tested at the 2-sided $\alpha$=0.04 level of significance and ORR will be tested at the 2-sided $\alpha$=0.01 level of significance.

All other statistical evaluations of efficacy will be considered exploratory.

## 7.4 Interim Analyses

The size of the trial is based upon the assumptions currently available and provides high power to detect what the Sponsor believes is the smallest clinically meaningful difference in OS under these assumptions. However, as there is uncertainty in the assumptions, interim analyses provides an opportunity to stop the trial early if the treatment benefit of cabozantinib is larger than expected, potentially allowing cabozantinib to become available sooner to this patient population.

Two interim analyses for primary endpoint of OS are planned. The interim analyses will be conducted at the time when approximately 311 and 466 deaths have been observed respectively. It is anticipated that this will be at approximately the 50% and 75% information fraction respectively for OS. Type I error for the interim analysis will be controlled by a Lan-DeMets O'Brien-Fleming alpha spending function as described in

Section 7.3. The actual critical values employed at the interim and final analyses of OS will depend upon the actual information fraction at the time of the interim analyses.

If the planned first or second interim OS analysis achieves overwhelming evidence of efficacy (p-value = 0.0031 or 0.0183, respectively, under trial design assumptions) in favor of cabozantinib, analyses of the secondary endpoints will be performed as planned. Interim analyses will be evaluated by the IDMC. Monitoring guidance will be provided to allow the study to be stopped early if the null hypothesis for OS is rejected in favor of cabozantinib. A nonbinding boundary for harm will be used by the IDMC to help evaluate subject safety. Stopping for futility is not planned.

## 7.5 Exploratory Efficacy Endpoints

Each exploratory endpoint will be analyzed using an appropriate two-sided statistical test without adjustment for multiplicity unless specified otherwise. Statistical results for exploratory endpoints will be considered descriptive. Exploratory analyses will be performed using all subjects in the ITT population unless specified otherwise.

### 7.5.1 Health-Related Quality of Life (HRQOL)

Health-related quality of life (HRQOL) will be assessed by the EuroQol Health questionnaire instrument (EQ-5D-5L). The questionnaire will be self-completed by the subjects at various time points until disease progression and will provide a generic measure of health for clinical appraisal (see protocol Section 5.5.8). The EQ-5D-5L questionnaire has two pages: a descriptive page which assesses on an increasing severity scale of 1-5 changes in the following five questions (dimensions): mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. The second page has a 0-100 visual analogue scale (VAS) which records the subjects self-rated health between 100 (best health you can imagine) and 0 (worst health you can imagine) and serves as quantitative measure of health by the subject (see protocol Appendix F and EQ-5D-5L User Guide 2015).

To compare the two treatment arms the following summaries are planned at each time point for each of the 6 questions:

Within each treatment arm:

- Descriptive statistics (number of observations, mean and standard deviation)
- Rate of completion for the questionnaire at each time point. This is defined as total number of subjects who answered all questions on the EQ-5D-5L questionnaire / the expected total number of subjects still on study at the visit

- Mean change from baseline at each time point and the corresponding 95% CI and p-value from one-sample t-test
- Effect size for change from baseline within arm, calculated as (mean of change in score/pooled standard deviation for baseline scores). An effect size greater than ≥ 0.3 will be considered potentially clinically meaningful
- Shift in the severity scale since baseline

Across treatment arms

- The difference in the effect sizes will be presented
- Difference in mean change from baseline for the questions at the time point will be evaluated by a two sample t-test
- Line plots for mean±standard error and the corresponding mean for change from baseline over time. Data from both treatment arms will be displayed on the same plot. In addition, these plots will also show the average state of the subjects at 3 landmark points, namely, around end of treatment, around progression and around 30 days post treatment follow-up for the two treatment arms
- Percentage of subjects in Level 1 (no problem) vs. Levels 2-5 (any problems) will be summarized over time
- Percentage of subjects with any problems (Level 2-5) will be compared between the treatment arms using a bar chart

The EQ-5D-5L may be converted into a single index (EQ-Index) value normalized across different countries where the index is validated. See Appendix D for conversion details. For EQ-VAS and EQ-Index, descriptive statistics for change from baseline at each time will be presented. Plots for mean±standard error and mean change from baseline±standard error over all time points for the two treatment arms will be generated. In addition, repeated-measures mixed-effects models will be used to explore treatment differences over time. These analyses will include the outcome variable of QOL score change from baseline. The predictors (fixed effects) will be the baseline scores, treatment arms, visit, and randomization strata described in Section 2.5. The individual subject nested within the planned treatment arm will be the random effect. All available data will be included for the analysis. The estimated least squares means for the two treatment arm and their difference, the p-values comparing the 2 treatment arms and the effect size will be presented. No adjustment will be made for multiple comparisons. An effect size of differences in the ≥ 0.3 range will be considered potentially clinically meaningful.

### 7.5.2 Duration of Objective Response

Duration of objective response is defined as the time from the first documentation of objective response by the investigator that is subsequently confirmed at a visit that is ≥ 28 days later to disease progression or death due to any cause.

Duration of response (months) = (earliest date of progressive disease or death due to any cause or censoring – date of first objective response + 1)/30.4375

Duration of objective response will be computed only among subjects who experience an objective response (confirmed CR or confirmed PR). Dates of progression and censoring will be determined as described for the secondary endpoint analysis of PFS (see section 7.2.1).

Duration of objective response will be analyzed using the same analysis method (Kaplan-Meier) as for the analysis on PFS (see Section 7.2.1.3).

The analyses for duration of response will be performed only if ORR >10%.

### 7.5.3 Time to Objective Response

Time to objective response is defined as the time from randomization to the first documentation of objective response by the investigator that is subsequently confirmed at a visit that is ≥ 28 days later to disease progression or death due to any cause.

Time to objective response (months) = (date of first objective response - date of randomization + 1)/30.4375

Time to objective response will be computed only among subjects who experience an objective response (confirmed CR or confirmed PR), and arithmetic methods (not Kaplan-Meier) will be used.

The analyses for time to response will be performed only if ORR >10%.

### 7.5.4 Alpha-fetoprotein (AFP)

For each scheduled post-baseline visit the AFP values at baseline and change from baseline will be summarized with standard descriptive statistics by treatment arm. Descriptive statistics for best/worst percent change since baseline will also be presented per arm using all available data. Waterfall plot will be presented by treatment arm for best percent change since baseline.

Best change is defined as the largest reduction (or smallest increase if no reduction) and worst change is defined as the largest increase (or smallest reduction if no increase). Appropriate transformations may be applied to normalize the data for presentation or analysis.

### 7.5.5 Serum Bone Markers

Serum bone markers (CTx, and BSAP) results at baseline, Week 3, Week 5 and Week 9 will be summarized by descriptive statistics. In addition, percent change from baseline at Week 3, Week 5 and Week 9, as well as best/worst percent change from baseline will be presented per arm using all available data.

Best change is defined as the largest reduction (or smallest increase if no reduction) and worst change is defined as the largest increase (or smallest reduction if no increase). These summaries will be provided for all subjects in the ITT population.

Appropriate transformations may be applied to normalize the data for presentation or analysis.

The best percent change from baseline for CTx and BSAP will be displayed graphically for each arm using waterfall plots. However, for CTx, subjects with baseline values that are below the limit of quantification (BLQ) will not be included in summaries or figures.

### 7.5.6 Child-Pugh Scores

Change from baseline in Child-Pugh categories will be summarized as a shift table at each time point through to end of treatment.

### 7.5.7 Health Care Resource Utilization

For this study the following health care resource utilization (HCRU) parameters collected during the study observation period will be summarized:

- Days of hospitalization due to SAEs
- Days in intensive care unit (ICU) due to SAEs
- Number of emergency room (ER) visits due to SAEs
- Number of surgeries

The summaries will include:

- Number and percentage of subjects in each category of HCRU
- Descriptive statistics for each HCR category amongst those subjects who utilized the respective resource
- Total number of days or visits as applicable for each HCRU
- Per person year summary for each HCRU

To calculate the per person year value for a subject for a HCRU parameter, the numerator is the sum of the days or visits for that subject for the parameter; and the denominator is defined as: (safety observation period – date of randomization + 1) / 365.25.

### 7.5.8 Pharmacokinetics (PK)

Pharmacokinetics analyses are outside the scope of this plan. A separate PK analysis plan and report will be provided.

## 7.6 Subgroups

The following subgroups based on baseline characteristics and stratification factors will be explored for primary and secondary efficacy endpoints

- Age category
  - <65 years
  - 65 to <75 years
  - 75 to <85 years
  - ≥85 years
- Sex

        - Male
        - Female
    - Race
        - Asian
        - Black or African American
        - White
        - Rest of the races reported/Not Reported
    - Geographic Regions 1
        - Asia (excluding Japan)
        - Europe/Australia/New Zealand
        - North America (Canada/USA)
        - Other
    - Geographic Regions 2
        - Asia
        - Other Region
    - ECOG Performance status at baseline:
        - 0
        - 1
        - Missing
    - Etiology of disease per stratification factors per IxRS:
        - HBV [with or without HCV]
        - HCV [without HBV]
        - Other
    - Current etiology of disease (per cancer history CRF):
        - HBV [without HCV] (Yes/No/Unk)
        - HCV [without HBV] (Yes/No/Unk)
        - HBV and HCV (Yes/No/Unk)
        - Alcoholism (Yes/No/Unk)
        - Nonalcoholic Steatohepatitis (NASH) (Yes/No/Unk)
        - Other (Yes/No/Unk)
    - Presence of extrahepatic spread of disease and/or macrovascular invasion (per IxRS):
        - Yes
        - No
    - Presence of extrahepatic spread of disease and/or macrovascular invasion (per cancer history CRF):
        - Yes

        - No
    - Visceral sites other than liver, bone, bone+visceral sites other than liver per tumor assessment CRFs per investigator. Visceral sites other than liver will include lung kidney, pancreas and other sites based upon a manual review of the reported sites after all data has been entered in the database
    - Prior systemic non-radiation anti-cancer therapy regimens for advanced HCC per subject per history of non-radiation anti-cancer therapy CRF (1 vs ≥ 2)
    - Prior receipt of PD-1/PD-L1 per history of non-radiation anti-cancer therapy CRF (Yes, No)
    - Prior receipt of Regorafenib per history of non-radiation anti-cancer therapy CRF (Yes, No)
    - Prior receipt of Lenvatinib per history of non-radiation anti-cancer therapy CRF (Yes, No)
    - Prior receipt of Tivantinib per history of non-radiation anti-cancer therapy CRF (Yes. No)
    - Prior receipt of Ramucirumab per history of non-radiation anti-cancer therapy CRF (Yes. No)
    - VEGF-A amplification in circulating tumor cells (Yes, No, Unknown)
    - AFP (<400 ng/mL, ≥ 400 ng/L) at baseline
    - Tumor MET status
        - High
        - Low or Negative
        - Unknown

# 8 SAFETY SUMMARIES

All safety analyses will be performed using all subjects in the Safety population. No formal statistical comparison between the two treatments arms is planned.

Safety and tolerability will be assessed by the incidence of treatment emergent-adverse events (TEAEs), changes in laboratory parameters and vital signs from baseline, and ECOG PS.

An Independent Data Monitoring Committee (IDMC) will monitor safety during the study on a regular basis. The committee will operate independently from the Sponsor and the clinical investigators.

The primary responsibility of the IDMC is to review the accumulating safety data on a regular and an ad hoc basis and make recommendations to the Sponsor regarding the continued conduct of the study. Safety data will be provided at regular intervals to the IDMC in the form of summary reports or data listings from the Sponsor (blinded) or its designated representative (unblinded).

Details regarding IDMC membership, schedule and format of meetings, format for presentation of data, access to interim data, method and timing of providing interim reports to the IDMC, and other issues relevant to committee operations are described in the IDMC charter.

The IDMC members will use their expertise, experience, and judgment to evaluate the safety data from the trial and to recommend to the Sponsor whether the trial should continue or be stopped early for safety. No formal statistical rules recommending early stopping for safety are planned.

## 8.1 Adverse Events

Adverse event terms recorded on the CRF will be mapped to preferred terms and system organ class using the Medical Dictionary for Regulatory Activities (MedDRA). The severity of AEs will be measured by CTCAE (Cancer Therapy Evaluation Program 2009) version 4 guidelines. The investigator will judge each event to be “not related” or “related” to study treatment.

A TEAE is defined as any event with an onset date on or after the date of the first dose of study drug or any ongoing event on the date of the first dose of study drug that worsens in severity after the date of the first dose of study drug.

Only TEAEs with an onset date through the end of the safety observation period (see section 4.5) will be tabulated in summary tables.

For the purpose of calculating treatment emergence and inclusion in summary tables, incomplete onset dates will be imputed as detailed in Appendix A.

For AE reporting, percentages ≥10% will generally be presented as integers, those <10%

will be presented with 1 decimal place (e.g. X.X%). Rounding rules are provided in Appendix B. The calculations of percentages will be based on original unrounded values.

An overall summary of adverse events will be provided with the number and percent of subjects who experienced the following types of events during the safety observation period (unless otherwise noted) in each treatment arm:

- Subjects with a TEAE
- Subjects with a Related TEAE
- Subjects with a Serious TEAE
- Subjects with a Serious Related TEAE at any time
- Subjects with a Worst-Grade 3 or 4 TEAE
- Subjects with a Worst-Grade 3 or 4 Related TEAE
- Subjects with a Worst-Grade 4 TEAE
- Subjects with a Worst-Grade 4 Related TEAE
- Subjects with a Grade 5 TEAE at any time
- Subjects with a Grade 5 TEAE
- Subjects with a Grade 5 TEAE through 30 days after the last dose of the study treatment
- Subjects with a Grade 5 AE judged not to be causally related to PD through 30 days of last dose of the study treatment
- Subjects with a Grade 5 AE >30 days after the last dose of the study treatment
- Subjects with a Grade 5 AE judged not to be causally related to PD >30 days of last dose of the study treatment
- Subjects with a Related Grade 5 TEAE at any time
- Subjects with a Related Grade 5 TEAE
- All subjects who died
- Subjects who died within 30 days after date of last dose of study treatment
- Subjects with a TEAE leading to Dose Modification
- Subjects with a TEAE leading to Dose Reduction
- Subjects with a TEAE leading to Dose Hold
- Subjects with TEAE leading to Treatment Discontinuation
    - TEAEs not related to disease progression
    - TEAEs related to disease progression

The following summaries of AEs will be provided:

| TEAE included | Row-levels (sorted by) | Columns |
|---|---|---|
| ***Subject Incidence by SOC, Preferred Term and Severity*** | | |
| All | SOC and PT<br>(MedDRA standard) | Worst severity:<br>All Grades, Grade 3/4, Grade 4, Grade 5 |
| Related | SOC and PT<br>(MedDRA standard) | Worst severity:<br>All Grades, Grade 3/4, Grade 4, Grade 5 |
| Serious | SOC and PT<br>(MedDRA standard) | Worst severity:<br>All Grades, Grade 3/4, Grade 4, Grade 5 |
| Related Serious | SOC and PT<br>(MedDRA standard) | Worst severity:<br>All Grades, Grade 3/4, Grade 4, Grade 5 |
| ***Subject Incidence by Preferred Term and Severity*** | | |
| All | PT<br>(descending frequency of Any Grade) | Worst severity:<br>All Grades, Grade 3/4, Grade 4, Grade 5 |
| Related | PT<br>(descending frequency of Any Grade) | Worst severity:<br>All Grades, Grade 3/4, Grade 4, Grade 5 |
| Leading to dose reduction | PT<br>(descending frequency of Any Grade) | Worst severity:<br>All Grades, Grade 3/4, Grade 4, Grade 5 |
| Leading to dose hold | PT<br>(descending frequency of Any Grade) | Worst severity:<br>All Grades, Grade 3/4, Grade 4, Grade 5 |
| Leading to dose modification | PT<br>(descending frequency of Any Grade) | Worst severity:<br>All Grades, Grade 3/4, Grade 4, Grade 5 |
| Not Related to Disease Progression and Leading to Study Treatment Discontinuation | PT<br>(descending frequency of Any Grade) | Worst severity:<br>All Grades, Grade 3/4, Grade 4, Grade 5 |
| Related to Disease Progression and Leading to Study Treatment Discontinuation | PT<br>(descending frequency of Any Grade) | Worst severity:<br>All Grades, Grade 3/4, Grade 4, Grade 5 |
| All | PT<br>(descending frequency of Grade 3/4) | Worst severity:<br>All Grades, Grade 3/4, Grade 4, Grade 5 |
| Related | PT<br>(descending frequency of Grade 3/4) | Worst severity:<br>All Grades, Grade 3/4, Grade 4, Grade 5 |
| Leading to dose reduction | PT<br>(descending frequency of Grade 3/4) | Worst severity:<br>All Grades, Grade 3/4, Grade 4, Grade 5 |
| Leading to dose hold | PT<br>(descending frequency of Grade 3/4) | Worst severity:<br>All Grades, Grade 3/4, Grade 4, Grade 5 |
| Leading to dose modification | PT<br>(descending frequency of Grade 3/4) | Worst severity:<br>All Grades, Grade 3/4, Grade 4, Grade 5 |
| Through 30 days after the last dose of study treatment | PT<br>(descending frequency of Grade 5) | Worst severity:<br>Grade 5 |
| Greater than 30 days after the last dose of study treatment | PT<br>(descending frequency of Grade 5) | Worst severity:<br>Grade 5 |
| AEs judged not to be causally related to PD through 30 days of last dose of study treatment | PT<br>(descending frequency of Grade 5) | Worst severity:<br>Grade 5 |
| AEs judged not to be causally related to PD > 30 days of last dose of study treatment | PT<br>(descending frequency of Grade 5) | Worst severity:<br>Grade 5 |
| ***Subject Incidence of AEs with Odds ratio, Relative Risk and Risk Difference*** | | |
| All | PT<br>(descending frequency of difference) | All events |
| ***Subject Incidence by Special Criteria*** | | |
| Events with an increase in the experimental arm of ≥5% (All Grades) or ≥2% (Grade 3/4) | SOC and PT<br>(SOC per MedDRA standard, PT within SOC by decreasing difference between arms for All Grades) | Worst severity:<br>All Grades, Grade 3/4, Grade 4, Grade 5 |
| Subject incidence of non-serious adverse event with an increase in the cabozantinib arm of ≥ 5% (Any Grade) | SOC and PT<br>(SOC per MedDRA standard, PT within SOC by decreasing difference between arms for All Grades) | Worst severity:<br>All Grades, Grade 3/4, Grade 4, Grade 5 |
| All | PT<br>(descending frequency of difference in percent between the two arms for All Grades) | Worst severity:<br>All Grades, Grade 3/4, Grade 4, Grade 5 |
| All | PT<br>(descending frequency of difference in percent between the two arms for | Worst severity:<br>All Grades, Grade 3/4, Grade 4, Grade 5 |

| TEAE included | Row-levels (sorted by) | Columns |
|---|---|---|
| | Grade 3/4) | |
| All AEs for Subjects with Macrovascular Invasion per Cancer History CRF | PT<br>(descending frequency of Gr3/4 in the cabo arm) | Worst severity:<br>All Grades, Grade 3/4, Grade 4, Grade 5 |
| All AEs for Subjects without Macrovascular Invasion per Cancer History CRF | PT<br>(descending frequency of Gr3/4 in the cabo arm) | Worst severity:<br>All Grades, Grade 3/4, Grade 4, Grade 5 |
| All SAEs for Subjects with Macrovascular Invasion per Cancer History CRF | PT<br>(descending frequency of Gr3/4 in the cabo arm) | Worst severity:<br>All Grades, Grade 3/4, Grade 4, Grade 5 |
| All SAEs for Subjects without Macrovascular Invasion per Cancer History CRF | PT<br>(descending frequency of Gr3/4 in the cabo arm) | Worst severity:<br>All Grades, Grade 3/4, Grade 4, Grade 5 |

The following data listings will also be provided with indicators for grade, relationship, seriousness, dose day of the event start/stop, days since last dose, action taken with study treatment:

- All AEs
- Grade 5 AEs
- Serious AEs other than death

## 8.2 Deaths

All reported subject deaths and whether death was causally associated with the disease under study (HCC) will be summarized by treatment arm for all subjects in the Safety population. The primary cause of death recorded on the CRF will be mapped to preferred term and system organ class using MedDRA. The coded terms will be merged with AE records to determine the relationship to study treatment.

Deaths will be summarized in 2 main categories as follows:

- Deaths within 30 days after the date of receipt of the last dose of study treatment
- Deaths greater than 30 days after the date of receipt of last dose of study treatment

Summary of primary cause of death will be tabulated under each category by causality to study disease and relationship to study drug.

All reported subject deaths will be listed.

## 8.3 Laboratory Assessments

### 8.3.1 Variables

The following treatment-emergent laboratory abnormalities will be summarized. A treatment emergent laboratory abnormality is defined as any laboratory abnormality with an onset date on or after the date of the first dose of study drug.

| Category | Abnormality | SDTM LBTESTCD | Grading System |
|---|---|---|---|
| **Hematology** | WBC increased<br>WBC decreased | WBC | CTCAE |
| | ANC decreased | NEUT | CTCAE |
| | Lymphocytes increased<br>Lymphocytes decreased | LYM | CTCAE |
| | Platelets decreased | PLAT | CTCAE |
| | Hemoglobin increased<br>Hemoglobin decreased | HGB | CTCAE |
| **Serum chemistry** | Albumin decreased | ALB | CTCAE |
| | ALP increased | ALP | CTCAE |
| | Amylase increased | AMYLASE | CTCAE |
| | ALT increased | ALT | CTCAE |
| | AST increased | AST | CTCAE |
| | Calcium, corr increased<br>Calcium, corr decreased | CACORR | CTCAE |
| | Calcium, ion increased<br>Calcium, ion decreased | CAION | CTCAE |
| | Creatinine increased | CREAT | CTCAE |
| | GGT increased | GGT | CTCAE |
| | Glucose increased<br>Glucose decreased | GLUC | CTCAE |
| | LDH increased | LDH | Sponsor |
| | Lipase increased | LIPASE | CTCAE |
| | Magnesium increased<br>Magnesium decreased | MG | CTCAE |
| | Phosphate decrease | PHOS | CTCAE |
| | Potassium increased<br>Potassium decreased | K | CTCAE |
| | Sodium increased<br>Sodium decreased | NA | CTCAE |
| | Total bilirubin increased | BILI | CTCAE |
| | Uric acid increased[2] | CYURIAC | CTCAE |
| Urine chemistry | UPCR increased | PROTCRT | Sponsor |
| Endocrinology[1] | Thyroid Stimulating Hormone increased<br>Thyroid Stimulating Hormone decreased | TSH | HLN |

[1] TSH is held in the SDTM "chemistry" laboratory category; will use HLN = high, low, normal classification based on normal range

[2] Uric acid increases will be graded only as Grade 1 or Grade 4. Grade 2 is not defined per CTCAE v4 and Grade 3 cannot be distinguished from Grade 1 based upon the result alone.

Sponsor-defined grades are to be applied to the following analytes:

LDH

- Grade 1 if >ULN to ≤ 2xULN
- Grade 2 if >2xULN to ≤ 3xULN
- Grade 3 if >3xULN

UPCR

- Grade 1 if ≥ 17.0 to ≤ 121.0 mg/mmol (≥0.15 to ≤1.0 mg/mg)
- Grade 2 if >121.0 to ≤ 396.0 mg/mmol (>1.0 to <3.5 mg/mg)
- Grade 3 if >396.0 mg/mmol (>3.5 mg/mg)

### 8.3.2 Analysis

All laboratory data parameters and visits will include flags for values above or below laboratory reference ranges. Toxicity grades will be assigned programmatically by applying the CTCAE v4 guidelines. Only results with assessment dates through the end of the safety observation period (see section 4.5) will be tabulated in summary tables.

Laboratory summaries will be presented in SI units. Continuous laboratory test results will be summarized by treatment group using descriptive statistics for actual values and for changes from baseline by scheduled visit. Box-Whiskers plots may also be presented at each scheduled visit (with visits shown on x-axis) for some laboratory parameters. For continuous laboratory test results which are below or above the quantification level, the imputed values as described in Appendix C will be used for deriving the grade and then summarized.

Tables summarizing the incidence of laboratory abnormalities by maximum post-baseline CTCAE grade overall and by baseline grade will be presented. In addition, the following summaries will also be presented:

A]Liver function abnormalities will be assessed as follows:

- Shift from baseline based on normal ranges
- Summaries of subjects meeting Hy's Law laboratory screening criteria as shown below:
  - >3× ULN (ALT or AST), >2× ULN Total Bilirubin, and <2× ULN ALP
  - >3× ULN (ALT or AST), >2× ULN Total Bilirubin, and ≥2× ULN ALP
- Sponsor-defined liver function test surveillance criterion:
  - >10× ULN (ALT or AST) and >2× ULN Total Bilirubin

B] For renal failure surveillance, a summary of subjects meeting renal failure laboratory screening criteria as shown below will be provided:

- Serum creatinine >= 3.0xULN and >= 2.0x baseline value or
- eGFR ≤ 50% of the baseline value or

- eGFR < 30 mL/min/1.73 $m^2$ and ≥ 25% reduction from the baseline value
  eGFR = 186 x (Creatinine in mmol per L / 88.4)$^{-1.154}$ x (Age)$^{-0.203}$ x (0.742 if female) x (1.210 if black) [from the UK CKD eGuide on the Renal Association website: http://egfrcalc.renal.org/]

C] Categorical summaries for baseline status for TSH and Free T4 will be presented. In addition categorical summaries for post-baseline TSH and Free T4 status among subjects with TSH and Free T4 in the normal range at baseline will also be presented.

For descriptive summaries for change from baseline analyses, only central lab results will be considered. For analyses of shift in grade from baseline or worst grade after baseline, all available results will be considered regardless of whether from the central or local lab.

| Summary Type | Include central lab results? | Include local lab results? |
|---|---|---|
| Subject-Incidence of Treatment Emergent Laboratory Abnormalities in Selected Laboratory Tests by CTCAE Grade | Y | Y |
| Change from Baseline in Laboratory Values | Y | N |
| Shift from Baseline in Laboratory Values by CTCAE Grade | Y | Y |
| Shift from Baseline in Laboratory Values by High/Low/Normal | Y | Y |
| Shift from Baseline in Laboratory Values by Sponsor-defined Grades | Y | Y |
| Subject-Incidence of Laboratory Abnormalities with a Between-Arm Difference of ≥5% (All Grades) or ≥2% (Grades 3-4) | Y | Y |

## 8.4 Vital Signs

### 8.4.1 Variables

The following vital signs will be summarized.

- Systolic blood pressure (mmHg)
- Diastolic blood pressure (mmHg)
- Weight

### 8.4.2 Analysis

Summary tables of vital signs and change from baseline for each study visit will be presented. Only results with assessment dates through the end of the safety observation period (see section 4.5) will be considered for the summaries.

Subject-incidence of clinically meaningful vital sign results as shown below will also be presented:

- Subjects who got worse since baseline and met the following blood pressure criteria on 2 or more visits (need not be consecutive) after first dose (JNC criteria were modified to include single measurement per time point when triplicate assessments were unavailable; JAMA 2003:289:2560):
    - Pre-hypertension: SBP 120-139 mmHg or DBP 80-89 mmHg
    - Stage 1: SBP 140-159 mmHg or DBP 90-99 mmHg
    - Stage 2: (SBP >= 160 mmHg and DBP <120) or DBP 100-119 mmHg
    - Stage 3: DBP >= 120 mmHg
- Proportion of subjects with weight loss ≥ 10% after first dose

## 8.5 ECOG Performance Status

For the purpose of evaluating safety, ECOG will be summarized as shift from baseline tables for, at minimum, the worst value recorded after the initiation of study treatment. Only results with assessment dates through the end of the safety observation period (see section 4.5) will be tabulated in summary tables.

Frequencies of ECOG worsening of ≥ +1 and +2 change from baseline to worst value after first dose will also be summarized.

## 8.6 Electrocardiogram (ECG)

Only results with assessment dates through the end of the safety observation period (see section 4.5) will be considered for summaries. The following categorical summaries will be presented per investigator and per independent review:

- number of subjects with triplicate average QTc > 500 ms after first dose
- number of subjects with increase in triplicate average QTc from baseline of >60ms
- number of subjects with increase in triplicate average QTc from baseline of >30ms after first dose

For the above summaries, the most-recent average value from triplicate measurements taken before first dose will be used as baseline. If no triplicate measurements were taken before first dose, the most-recent single value taken before first dose will be used as baseline.

# 9 IDENTIFICATION AND SUMMARY OF PROTOCOL DEVIATIONS

In accordance with ICH E3, important eligibility deviations per the inclusion/exclusion criteria and important post-randomization protocol deviations tracked in CTMS (clinical trial management system) will be identified and listed separately by study center and subject. Important deviations will be summarized for the ITT population as follows:

Deviation code:

- DID NOT SATISFY INCLUSION OR EXCLUSION CRITERIA
- PROHIBITED MEDICATION
- TREATMENT DEVIATION
- WITHDRAWAL DEVIATION
- RANDOMIZATION IRREGULARITY
- OTHER PROTOCOL DEVIATION

Deviations category:

- IMPORTANT
- OTHER

Deviations sub-category:

- POTENTIALLY IMPACTING SAFETY
- POTENTIALLY IMPACTING EFFICACY
- POTENTIALLY IMPACTING SAFETY AND EFFICACY
- OTHER

# 10 DATA QUALITY ASSURANCE

The Clinical, Data Management, Biostatistics, and Medical Writing departments at Exelixis and designees will work diligently and collaboratively to ensure that the data collected and analyzed for this study are of the high quality. In addition to electronic evaluation of the data and verification of data from source documents at the respective sites, a data review meeting will be held to review the data and correct significant data anomalies before the study database is locked or data are extracted for the purpose of analysis.

# 11 REFERENCES

Chobanian AV, Bakris GL, et al. The Seventh Report of the Joint National Committee On Prevention, Detection, Evaluation, And Treatment of High Blood Pressure:The JNC 7 Report. JAMA 2003 May 21; 289(19):2560-72.

Common Terminology Criteria for Adverse Events (CTCAE) Version 4.0, USDHHS, NIH, NCI; publish date May 28, 2009 (v4.03: June 14, 2010).

International Conference on Harmonization ICH E9: Statistical principles for clinical trials (05 February 1998 [Europe], September 1998 [FDA]).

Llovet JM, Decaens T, Raoul J-L, et al. Brivanib versus placebo in patients with advanced hepatocellular carcinoma (HCC) who failed or were intolerant to sorafenib: results from the Phase 3 BRISK-PS study. J Hepatol 2012;56 Suppl 2, Abstract 1398.

Reenen MV, Janssen B et al. EQ-5D-5L User Guide, Version 2.1, April 2015

## Appendix A: Date Imputation Rules

Incomplete Cancer Diagnosis Date

If *year* is missing (or completely missing): do not impute

If only *day* is missing: set to 15th of the month.

If *day* and *month* are missing: set to July 1st.

If either imputation rule above results in a diagnosis date > informed consent:

set diagnosis date to the date of informed consent - 1.

Incomplete Adverse Event Onset Date

Assumption: For on-study Adverse Events.

If *year* is missing (or completely missing): set to the date of first dose.

If (*year* is present and *month* and *day* are missing) or (*year* and *day* are present and *month* is missing):

If *year* = year of first dose: set the date to the first dose date.

If *year* < year of first dose: set *month* and *day* to December 31st.

If *year* > year of first dose: set *month* and *day* to January 1st.

If *month* and *year* are present and *day* is missing:

If *year* = year of first dose, and:

If *month* = month of first dose: set *day* to day of first dose.

If *month* < month of first dose: set *day* to last day of *month*.

If *month* > month of first dose: set *day* to 1st day of *month*.

If *year* < year of first dose: set *day* to last day of month.

If *year* > year of first dose: set *day* to 1st day of month.

For all other cases: set to date of first dose.

Incomplete Concomitant Medication Start Date

If *year* is missing (or completely missing): do not impute.

If (*year* is present and *month* and *day* are missing) or (*year* and *day* are present and *month* is missing):

Set *month* and day to January 1st.

If *year* and *month* are present and *day* is missing:

Set *day* to 1st day of month.

Incomplete Concomitant Medication End Date

If *year* is missing (or completely missing): do not impute.

If (*year* is present and *month* and *day* are missing) or (*year* and *day* are present and *month* is missing):

Set *month* and day to December 31st.

If *year* and *month* are present and *day* is missing:

Set *day* to last day of the month.

Incomplete Subsequent Anti-Cancer Therapy Start Date

Assumption: Anti-Cancer therapies reported on the Subsequent Anti-Cancer Therapy CRF.

If *year* is missing (or completely missing): set to date of last dose of study treatment + 1

If (*year* is present and *month* and *day* are missing) or (*year* and *day* are present and *month* is missing):

If *year* > year of the last dose: Set *month* and *day* to January 1st.

If *year* = year of the last dose: Set *month* and *day* to date of last dose of study treatment + 1

If *year* and *month* are present and *day* is missing:

Set *day* to 1st day of month if the resulting imputed date is greater than date of last dose or if the month is before the month of last dose date and year is same or before the year of the last dose date. Otherwise set the imputed date to date of last dose + 1

Incomplete Death Date

Identify date of last known alive (LA) prior to death from the following:

1. Date of decision to discontinue study treatment from End of Treatment CRF
2. Date of last radiographic assessment from End of Radiographic Follow Up CRF
3. Date last known alive from Survival Follow Up CRF
4. Date of last lab assessment from the Labs dataset

If *year* is missing (or completely missing): set to date of LA + 1

If only *day* is missing: set to the maximum of the first of month or LA + 1

If *month* and *day* are missing:

If *year* of LA = year of death

Set death date to date of LA + 1

If *year* of most-recent contact < year of death

Set *month* and *day* to Jan 1st.

Incomplete Study Treatment Start Date

Define previous sequential dosing "milestone" as the latest of previous dose stop date, previous dose hold stop date, date of first dose or randomization date.

If *year* is missing (or completely missing): set to date of previous sequential dosing "milestone" + 1

If (*year* is present and *month* and *day* are missing) or (*year* and *day* are present and *month* is missing): set to January 1st

If *year* and *month* are present and *day* is missing: set to the first day of the month

If the imputed date is before the previous sequential dosing "milestone": set to the date of previous sequential dosing "milestone" + 1

Incomplete Study Treatment Stop Date

Define next sequential dosing "milestone" as the earliest of next dose start date, next dose hold start date, date of last dose from EOT CRF or the cutoff date.

If *year* is missing (or completely missing): set to date of next sequential dosing "milestone" - 1

If (*year* is present and *month* and *day* are missing) or (*year* and *day* are present and *month* is missing): set to December 31st

If *year* and *month* are present and *day* is missing: set to the last day of the month

If the imputed date is after the next sequential dosing "milestone": set to the date of next sequential dosing "milestone" - 1

## Appendix B: Rounding Rules for Reported Percentages

For percentages ≥10%:

- Values >= X.5 or above round to X+1.
- Values >X but <X.5 round to X.

For percentages <10%:

- Values >= X.Y5 or above round to X.Y+0.1.
- Values >X.Y but <X.Y5 round to X.Y.

## Appendix C: Imputation Rules for Laboratory Values Outside of Quantification Range

- Lab values below the lower level of quantification (LLQ) that are reported as "<LLQ" or "≤LLQ" in the database will be imputed by LLQ x 0.99 for analysis purposes. However the original value will also be maintained.

- Lab values above the upper level of quantification (ULQ) that are reported as ">ULQ" or "≥ULQ" in the database will be imputed by ULQ x 1.01 for analysis purposes. However the original value will also be maintained.

## Appendix D: EQ-5D-5L Index Value Conversion Guidelines

The EQ-index conversion algorithm (EQ-5D-5L User Guide 2.1, April 2015. Available from: http://www.euroqol.org/about-eq-5d/publications/user-guide.html):

- Calculate ***health state***
  - Each of the 5 dimensions comprising the EQ-5D descriptive system is divided into 5 levels of perceived problems:
    - Level 1: indicating no problem
    - Level 2: indicating slight problems
    - Level 3: indicating moderate problems
    - Level 4: indicating severe problems
    - Level 5: indicating extreme problems
  - A unique health state is defined by combining 1 level from each of the 5 dimensions.
    For example, state 11111 indicates no problems on any of the 5 dimensions, while state 12345 indicates no problems with mobility, slight problems with washing or dressing, moderate problems with doing usual activities, severe pain or discomfort and extreme anxiety or depression.
    Note that missing values will be coded as '9'. Ambiguous values will be treated as missing values.
- ***EQ-index values*** for each country = ***health state * the country specific conversion factors*** for each dimension (EQ-5D-5L Index Value Calculator, version 1)