I3O-MC-JSBF Statistical Analysis Plan Version 2

Randomized, Double-Blind, Phase 2 Study of Ramucirumab or Merestinib or Placebo plus Cisplatin and Gemcitabine as First-Line Treatment in Patients with Advanced or Metastatic Biliary Tract Cancer

NCT02711553

Approval Date: 19-Mar-2018

1. Statistical Analysis Plan: I3O-MC-JSBF: Randomized, Double-Blind, Phase 2 Study of Ramucirumab or Merestinib or Placebo plus Cisplatin and Gemcitabine as First-Line Treatment in Patients with Advanced or Metastatic Biliary Tract Cancer

Merestinib (LY2801653) Ramucirumab (LY3009806) Biliary Tract Cancer

Eli Lilly and Company
Indianapolis, Indiana USA 46285

Protocol I3O-MC-JSBF
Phase 2

Statistical Analysis Plan Version 1 electronically signed and approved by Lilly on 08 September 2016
Statistical Analysis Plan Version 2 electronically signed and approved by Lilly on date provided below

Approval Date: 19-Mar-2018 GMT

2. Table of Contents

Section **Page**

Table of Contents

Table **Page**

Table of Contents

Figure **Page**

3. Revision History

SAP Version 1 was approved prior to unblinding.

SAP Version 2 was updated and approved prior to the primary analysis of PFS.

The overall changes and rationales for the changes incorporated in Version 2 are as follows:

- Added that an interim efficacy analysis of OS will be conducted at the time of primary analysis of PFS, and specified the statistical assumptions for the interim analysis.
- Revised the unblinding plan to specify that (1) only assessment committee (AC) will assess primary analysis of PFS and interim efficacy analysis of OS, and (2) the roles of limited sponsor personnel outside of AC who may be unblinded after primary PFS analysis to perform analyses and make critical decisions.
- Updated the analysis of OS that the overall Type I error rate for statistical comparison between each investigational treatment to pooled control will be controlled at 1-sided 0.1, and that for regulatory purposes, the same comparison will also be conducted using 1-sided alpha level of 0.025.
- Specified that gatekeeping approach will be used for the testing of PFS and OS.

4. Study Objectives

4.1. Primary Objective

Evaluate the efficacy, in terms of progression-free survival (PFS) in patients with advanced or metastatic biliary tract cancer (BTC), for

- ramucirumab (8 mg/kg, IV) versus placebo, in combination with cisplatin (25 mg/m^2, intravenous [IV]) and gemcitabine (1000 mg/m^2, IV), on Days 1 and 8 of a 21-day cycle
- merestinib (80 mg, oral each day) versus placebo, in combination with cisplatin (25 mg/m^2, IV) and gemcitabine (1000 mg/m^2, IV), on Days 1 and 8 of a 21-day cycle

4.2. Secondary Objectives

- Evaluate the efficacy, in combination with cisplatin and gemcitabine, in terms of overall survival (OS), for ramucirumab versus placebo and merestinib versus placebo
- Evaluate the efficacy, in combination with cisplatin and gemcitabine, in terms of objective response rate (ORR) and disease control rate (DCR) for ramucirumab versus placebo and merestinib versus placebo
- Safety profile of ramucirumab; safety profile of merestinib
- Pharmacokinetics (PK) of ramucirumab; PK of merestinib
- Immunogenicity of ramucirumab
- Evaluate patient-reported outcomes (PRO) measures of disease-specific symptoms, in combination with cisplatin and gemcitabine, for ramucirumab versus placebo and merestinib versus placebo.

4.3. Exploratory Objectives

- To explore biomarkers relevant to ramucirumab or merestinib, angiogenesis, immune function, and the disease state, and to correlate these markers to clinical outcome.
- Explore the relationship of pretreatment weight loss with primary and secondary treatment response outcomes.

5. Study Design

5.1. Summary of Study Design

Study I3O-MC-JSBF (JSBF) is a global, multicenter, randomized, double-blind, Phase 2 study in patients with advanced or metastatic BTC who have had no prior therapy for advanced or metastatic BTC.

The primary objective is to compare the PFS in BTC patients treated with ramucirumab plus cisplatin and gemcitabine versus placebo plus cisplatin and gemcitabine and merestinib plus cisplatin and gemcitabine versus placebo plus cisplatin and gemcitabine. To assess the primary objective, approximately 300 patients will be randomized to receive IV/oral investigational treatment or IV/oral control in a 2:1:2:1 fashion:

Intravenous Treatment:

- Arm A1: ramucirumab 8 mg/kg plus cisplatin (25 mg/m^2) and gemcitabine (1000 mg/m^2) intravenously on Days 1 and 8, every 21 days

or

- Arm A2: placebo plus cisplatin (25 mg/m^2) and gemcitabine (1000 mg/m^2) intravenously on Days 1 and 8, every 21 days

Oral Treatment:

- Arm B1: merestinib 80 mg orally each day, plus cisplatin (25 mg/m^2, IV) and gemcitabine (1000 mg/m^2, IV) on Days 1 and 8, every 21 days

or

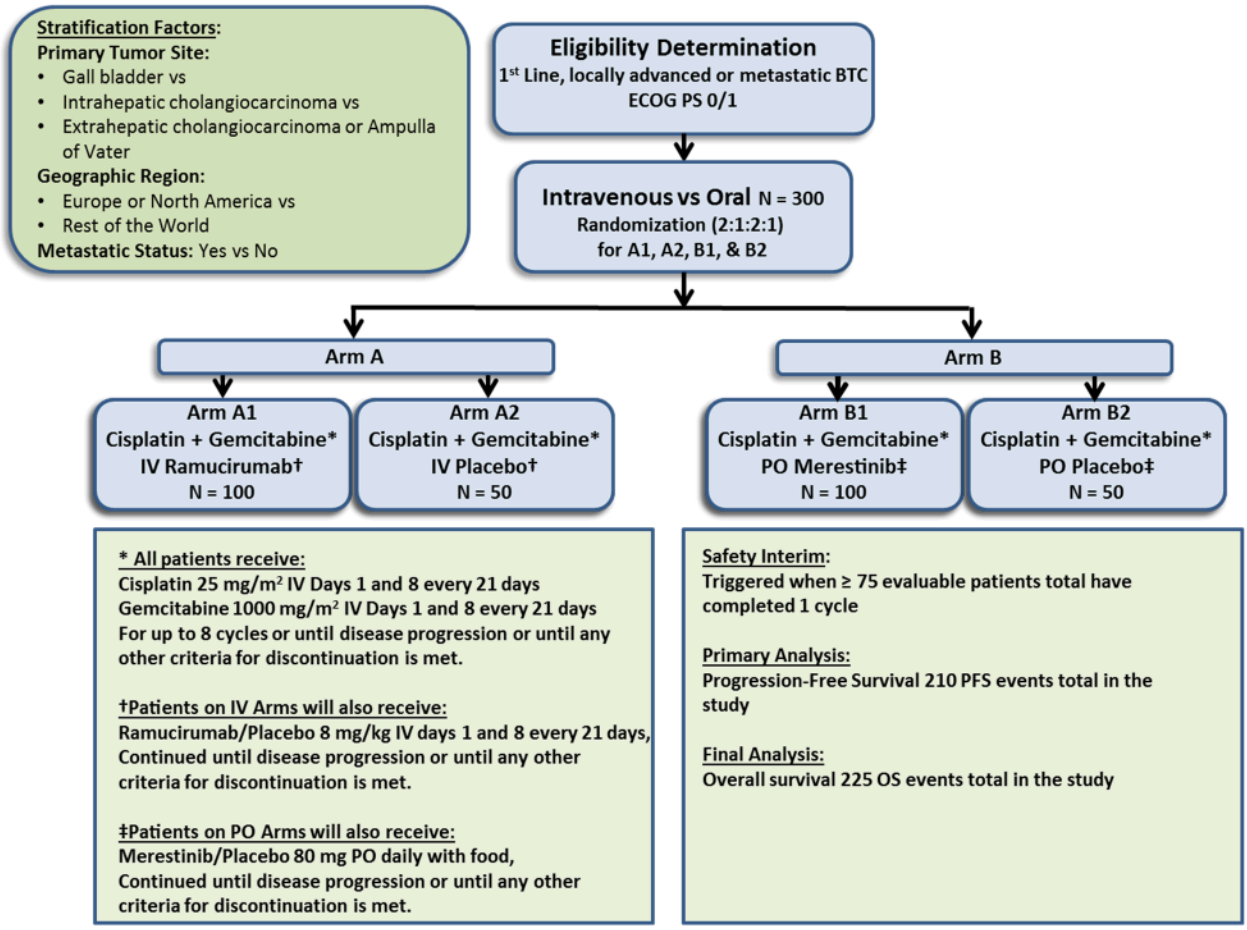
- Arm B2: placebo orally each day, plus cisplatin (25 mg/m^2, IV) and gemcitabine (1000 mg/m^2, IV) intravenously on Days 1 and 8, every 21 days

A cycle is defined as an interval of 21 days. Treatment with all study drugs will be given in an outpatient setting. Treatment will continue until there is evidence of disease progression or any other discontinuation criteria are met. Treatment with cisplatin and gemcitabine will be capped at a maximum of 8 cycles. Treatment with the randomly assigned study therapy of ramucirumab, merestinib, or placebo will not be capped at a maximum number of cycles and should be continued until there is evidence of disease progression or any other discontinuation criteria are met. During the planned treatment period, if one or more therapeutic agent is permanently discontinued for a reason other than progressive disease, then treatment with the other study agent(s) should continue and the patient should remain on study with full adherence to all protocol-related requirements.

Crossover between arms is not permitted.

All patients will be offered best supportive care, as determined appropriate by the investigator and as not otherwise limited specifically within the protocol.

Figure 5.1 illustrates the study design.



Abbreviations: ECOG PS = Eastern Cooperative Oncology Group performance status; IV = intravenous; OS = overall survival; PFS = progression-free survival; PO = oral.

Figure 5.1 Illustration of study design

5.2. Determination of Sample Size

The study will randomize 300 patients. The primary analysis of PFS for the comparison of each investigational treatment to pooled control will be performed when a minimum of a total of 210 events in the study have been observed, assuming a 30% censoring rate in the study.

Assuming a PFS hazard ratio (HR) of 0.70, this sample size yields approximately 80% statistical power to detect superiority of each investigational treatment over the pooled control with the use of a log-rank test, and a 1-sided type I error of 0.10, taking into account of the worst case scenario that the same numbers of events are observed in an investigational treatment and pooled control.

A gatekeeping approach will be used to assess PFS and OS. The OS endpoint will only be tested if the PFS test is significant, to control overall Type I error across both endpoints. There are two opportunities for testing the OS endpoint – at the primary PFS analysis (as long as the PFS test is

significant) as well as the final OS analysis, which will be performed for the comparison of each investigational treatment to pooled control when a minimum total of 225 OS events in the study have been observed.

Assuming an OS HR of 0.75 and an interim efficacy analysis of OS to be conducted at primary analysis of PFS with approximately 50% of information accumulated based on projection, this sample size yields approximately 68% statistical power to detect superiority of each investigational treatment over the pooled control with the use of a stratified log-rank test and a 1-sided type I error rate of 0.10. The Lan-DeMets (O'Brien-Fleming) alpha-spending function will be used to control the overall Type I error rate for each comparison. The nominal significance level for each comparison is 0.02 at interim efficacy analysis and 0.094 at the final OS analysis. The actual type I error rate spent at interim efficacy analysis will be adjusted based on the actual number of events observed when the interim efficacy analysis of OS takes place, if it is materially different from the projection.

5.3. Method of Assignment to Treatment

Approximately 300 patients will be randomly assigned. Patients who meet all criteria for enrollment will be randomly assigned within 24 to 72 hours prior to Visit 1 (Day 1, Cycle 1). Every attempt should be made to randomize the patient as close as possible to Day 1 of Cycle 1 and not more than 72 hours prior to Day 1. Assignment to treatment groups will be determined by a computer-generated random sequence using an interactive Web response system (IWRS).

Randomization will be stratified by the following factors:

- Primary Tumor Site (gall bladder versus intrahepatic cholangiocarcinoma versus extrahepatic cholangiocarcinoma or Ampulla of Vater)
- Geographic regions (North America or Europe versus Rest of World)
- metastatic disease (yes versus no)

The chosen stratification factors have been identified as potentially important adverse prognostic factors or regional considerations in this disease. Randomization will be performed separately within each of the strata (or cells) defined by all combinations of these 3 variables.

6. A Priori Statistical Methods

6.1. General Considerations

Statistical analysis of this study will be the responsibility of Eli Lilly and Company. The interpretation of the study results will be the responsibility of the Lilly Clinical Research Physician/Clinical Research Scientist (CRP/CRS) and statistician. The CRP/CRS and statistician will also be responsible for the appropriate conduct of internal reviews for both the final study report and any study-related material to be authorized by Lilly for publication.

Any change to the data analysis methods described in the protocol will require an amendment ONLY if it changes a principal feature of the protocol. Any other change to the data analysis methods described in the protocol and the justification for making the change will be described in the clinical study report.

The following populations will be analyzed in this study:

- **Intention-to-Treat (ITT) population:** will include all randomized patients. The ITT analysis of efficacy data will consider allocation of patients to treatment groups as randomized and not by actual treatment received. This population will be used for all baseline, efficacy, and health economics analyses.
- **Per-protocol population:** will include all randomized patients who receive at least 1 dose of study treatment and do not have any major protocol violations that could potentially affect the efficacy conclusions of the study. This population will be determined after study team has reviewed the list of important protocol deviations at end of the study, following the guideline below:
    - The absence of the violation of inclusion/exclusion criteria, significant deviations in efficacy assessments, and instances of suspected misconduct
- **Safety population:** will include all randomized patients who received any quantity of study treatment, regardless of their eligibility for the study. The safety evaluation will be performed based on the first dose of study treatment a patient actually received, regardless of the arm to which he or she was randomized. The safety population will be used for all dosing/exposure, safety, and resource utilization analyses.
- **Pharmacokinetic population:** will include all treated patients who received at least 1 dose of study treatment and have at least 1 post baseline evaluable PK sample.
- **Biomarker population:** will include the subset of patients from the randomized population from whom a valid assay result has been obtained.

All tests of treatment effects will be conducted at a 1-sided alpha level of .1, unless otherwise stated. All confidence intervals (CIs) will be given at a 2-sided 95% level, unless otherwise stated.

Summary statistics for continuous variables will include number of patients (N), mean, median, standard deviation (SD), minimum, and maximum. Summary statistics for categorical endpoints

will include N, frequency, percentages, and associated standard error (SE) and 95% CI. Exploratory analyses will be conducted as deemed appropriate. Time-to-event variables will be summarized using the Kaplan-Meier (KM) method. Kaplan-Meier estimates of median and quartiles will be reported along with 95% CIs.

6.2. Handling of Dropouts or Missing Data

Baseline will refer to the last non-missing observation prior to first administration of any treatment unless stated otherwise. Missing data, except for dates, will not be imputed. When dates are used in calculations, missing days will be replaced with 15th of the month and missing day/month with 01 JULY. Where windows are allowed for data collection and there is more than one reading in any window, appropriate consideration will be given as to whether only one value from the window should be used, and if so how it should be chosen. This could either be the mean (geometric mean) or the value closest to the mid-point of the window or the value closest to the data collection time of another variable if the analysis involves time-matched analyses.

If a patient discontinues from study treatment for any reason other than becoming lost to follow-up or death, regular follow-up visits will be conducted to collect information on subsequent non-protocol defined therapy, first disease recurrence, and death. Therefore, it is possible to obtain information for time-to-event endpoints for these patients.

In the efficacy analysis, a patient with a tumor response of inevaluable will be included in the denominator for the purpose of calculating ORR.

6.3. Multiple Comparisons/Multiplicity

The Internal Assessment Committee (IAC) will perform unblinded interim efficacy analysis of OS. The group sequential design with gatekeeping approach based on the Lan-DeMets (O'Brien-Fleming) alpha-spending function controls the type I error rate for the key efficacy endpoints including PFS and OS. The interim analysis of safety do not affect the type I error rate.

6.4. Patient Disposition

A detailed description of patient disposition will be provided. It will include a summary of the number and percentage of patients randomized in the study, reasons for discontinuation from study treatment, and reasons for discontinuation from the study. Reason for discontinuation from both study treatment and the study will be listed by the pre-determined categories. For treatment discontinuation, these include progressive disease, adverse event (AE), death, withdrawal by subject, physician decision, non-compliance with study drug, protocol deviation, study terminated by IRB/ERB/sponsor, lost to follow-up; for study discontinuation, these include study terminated by sponsor, withdrawal by subject, lost to follow-up, death. If the reason for discontinuation is AE or death, the associated AE or cause of death will be reported. The disposition will also be listed. All patients randomized in the study will be included in the summaries and listings.

6.5. Patient Characteristics

Patient characteristics will be summarized and listed for all patients randomized by treatment, including:

- Patient demographics (including age, sex, race and ethnicity, screening height and weight, and screening derived body surface area)
- Baseline disease characteristics (including basis for initial diagnosis, initial pathological diagnosis, stage at initial diagnosis, baseline Eastern Cooperative Oncology Group performance status)

6.6. Treatment Compliance

Ramucirumab, IV placebo, cisplatin, and gemcitabine will be administered only at the investigational sites by the authorized study site personnel. As a result, treatment compliance is ensured.

Patient compliance with merestinib and oral placebo will be assessed at each visit by direct questioning and counting returned tablets. Dose compliance will be summarized for each oral treatment. Dose compliance will be calculated as (total amount of drug taken (units)/total amount of drug prescribed (units))*100%. "Total amount of drug taken" will be derived from the difference between the total number of tablets dispensed and returned over the course of the patient's treatment across all visits prior to treatment discontinuation. "Total amount of drug prescribed" is the sum of products of dosing intervals and the expected dose for each interval. Total dose prescribed should take into consideration any dose adjustment(s) before treatment discontinuation date.

6.7. Concomitant Therapy

Prior and concomitant medications and therapies will be listed and summarized by treatment arm.

The numbers and percentages of patients receiving post-discontinuation anticancer therapies will be provided by type of therapy (surgery, radiotherapy, or systemic therapy), and by drug class and/or name, overall and by line of therapy.

6.8. Efficacy Analyses

6.8.1. Primary Outcome and Methodology

Progression-free survival is defined as the time from randomization until the first investigator-determined objective progression as defined by Response Evaluation Criteria In Solid Tumors Version 1.1 (RECIST v1.1), or death from any cause in the absence of progressive disease. Patients known to be alive and without disease progression will be censored at the time of the last adequate tumor assessment (a detailed PFS event/censoring scheme is provided in Table JSBF.6.1).

Progression-free survival will be compared between each investigational treatment to pooled control using a stratified log-rank test and 1-sided alpha level of 0.1, stratified by the randomization strata. For regulatory purposes, the same comparison will also be conducted using 1-sided alpha level of 0.025. The corresponding HR between treatment arms will be estimated using a stratified Cox regression model (Cox 1972), stratified by the randomization strata. Progression-free survival curves, median PFS, and PFS rates at 6 and 12 months with 80% and 95% CIs for each treatment arm will be estimated using the Kaplan-Meier method (Kaplan and Meier 1958).

Table JSBF.6.1. PFS Event/Censoring Scheme

Situation[a]	Event/Censor	Date of Event or Censor
Investigator assessed tumor progression or death	Event	Earliest date of PD or death
No tumor progression and no death	Censored	Date of last adequate radiological assessment or date of randomization (whichever is later)[b]
Unless		
No baseline radiological tumor assessment available	Censored	Date of randomization
No adequate postbaseline radiological tumor assessment available and death reported after 2 scan intervals following randomization[b,c]	Censored	Date of randomization
New anticancer treatment started and no tumor progression or death within 14 days	Censored	Date of adequate radiological assessment prior to (start of new therapy +14 days) or date of randomization (whichever is later)
Tumor progression or death documented immediately after 2 or more scan intervals following last adequate radiological tumor assessment or randomization (whichever is later)[b,c]	Censored	Date of last adequate radiological assessment or date of randomization (whichever is later)[b]

Abbreviations: CR = complete response; PD = progressive disease; PFS = progression-free survival; PR = partial response; SD = stable disease.

[a] Symptomatic deterioration (that is, symptomatic progression that is not radiologically confirmed) will not be considered as tumor progression.

[b] Adequate radiological tumor assessment refers to an assessment with one of the following responses: CR, PR, SD, or PD.

[c] Radiologic imaging for tumor assessment will be performed every 6 weeks (±7 days) after randomization (regardless of treatment delays) during the Study Treatment Period, until disease progression OR study completion or 14 months after randomization, whichever occurs first. Refer to footnote b for any patient whose disease has not progressed by 14 months after randomization. Any patient whose disease has not progressed by 14 months after randomization will be evaluated for response every 12 weeks (±7 days) from 14 months after randomization, until disease progression OR study completion OR 3 years after randomization, whichever occurs first; then as per standard clinical practice after that.

6.8.2. Secondary Efficacy Analyses

Overall survival is defined as the time from randomization until death from any cause. If the patient is alive or lost to follow-up at the time of data analysis, OS data will be censored on the last date the patient is known to be alive. Using gatekeeping approach, overall survival will be compared between each investigational treatment to pooled control using a stratified log-rank test and 1-sided overall alpha level of 0.1, stratified by the randomization strata. For regulatory purposes, the same comparison will also be conducted using 1-sided alpha level of 0.025. The corresponding HR between treatment arms will be estimated using a stratified Cox regression model (Cox 1972), stratified by the randomization strata. Overall survival curves, the median and survival rates at 6 and 12 months with 95% CI, for each treatment arm will be estimated using the Kaplan-Meier method (Kaplan and Meier 1958).

Overall response rate is defined as the number of patients who achieve a best overall response of CR or PR divided by the total number of patients randomized to the corresponding treatment arm (ITT population). The confirmation of CR and PR is not required. The ORR, with 95% CI, will be summarized for each treatment arm and compared between each investigational treatment to pooled control using the Cochran-Mantel-Haenszel test adjusting for the randomization strata.

Disease control rate (DCR) is defined as the number of patients who achieve a best overall response of CR, PR, or SD divided by the total number of patients randomized to the corresponding treatment arm (ITT population). The confirmation of CR and PR is not required. In the case of SD, measurements must have met the SD criteria at least once after study entry at a minimum interval (in general not less than 6 to 8 weeks). The DCR, with 95% CI, will be summarized for each treatment arm and compared between each investigational treatment to pooled control using the Cochran-Mantel-Haenszel test adjusting for the randomization strata.

Duration of response (DoR) is defined as the time from the date measurement criteria for CR or PR (whichever is first recorded) are first met until the first date that disease is recurrent or objective progression is observed, per RECIST 1.1 criteria, or the date of death from any cause in the absence of objectively determined disease progression or recurrence. Duration of response will be summarized for each treatment arm using descriptive statistics.

6.8.3. Sensitivity Analyses

Unstratified analyses of aforementioned primary and secondary efficacy outcomes may be conducted.

6.8.3.1. Bayesian sensitivity analyses for PFS

Sensitivity of comparative inference for PFS based on the shared control (i.e. pooling patients from study arms A2 and B2) may be accomplished via Bayesian mixture modeling as follows. For now, we may assume a constant hazards model (event times are thereby presumed to be distributed exponential), but note that extension to a piecewise exponential representation may be adopted upon determination of underlying lack of fit for the simpler model. Both arms A1 and B1 will be simultaneously compared to control therapy (i.e. within the same model). Let λ_{A2} denote hazard for patients in study arm A2 and, likewise, let λ_{B2} denote hazard in study arm B2. The following hierarchical model facilitates *adaptive pooling* of data from A2 and B2 while simultaneously informing the comparison of experimental treatment to control.

$$e_i \sim Poisson(t_i \cdot \theta_i), \text{ (Independently over } i)$$

$$\theta_i = \begin{cases} \lambda_{A2}, & i \in A2 \\ \lambda_{B2}, & i \in B2 \\ \lambda_{A2} \cdot \exp(\beta_{A1}), & i \in A1 \\ \lambda_{B2} \cdot \exp(\beta_{B1}), & i \in B1 \end{cases}$$

$$\lambda_{A2} \sim Gamma(\alpha, \beta)$$

$$\lambda_{B2} \sim Gamma(\alpha, \beta)$$

$$\lambda_{B2} | \lambda_{A2} \sim \begin{cases} \delta_{\lambda_{A2}}(\lambda_{B2}), & w.p. \quad \rho \\ Gamma(\alpha, \beta), & w.p. \quad (1-\rho) \end{cases}$$

$$\beta_{A1} \sim Normal(\mu, \tau^2)$$

$$\beta_{B1} \sim Normal(\mu, \tau^2)$$

The idea is that with prior probability ρ (which the statistician must specify), the control arms A2 and B2 are pooled to inform a single control hazard rate, say $\lambda = \lambda_{A2} = \lambda_{B2}$. With prior probability $(1-\rho)$, however, patients from arms A2 and B2 are not pooled, and inform *separate* arm-specific control hazard rates (so that $\lambda_{A2} \neq \lambda_{B2}$). The goal of this is to acknowledge that there may, in fact, be some underlying difference in hazard rates from study arms A2 and B2 (even though the therapy given is quite similar) and to let the data drive the degree to which information is pooled. If there is clear evidence of dissimilar hazard in these two arms, the *posterior* probability of pooling (which will update away from ρ) may decrease to favor distinct values $\lambda_{A2} \neq \lambda_{B2}$. On the other hand, if there isn't statistically meaningful evidence of a difference, the model may favor pooling. Ultimately, all uncertainty (regarding the degree of difference between the control arms, and the treatment vs. control comparisons themselves) will be captured within this single model and may provide additional statistical context to primary inference. All hyperparameters will be selected so as to inject minimal influence on quantities of interest (particularly the log hazard ratios β_{A1} and β_{B1}). The shape and scale parameters α and β

must be selected so as to yield relatively non-informative priors for the hazard rate, but *not* overly diffuse priors, as the diffusivity of this prior may unnaturally force posterior probability of pooling to zero.

6.9. Health Outcomes/Quality-of-Life Analyses

The main analysis will be conducted in the ITT population. Exploratory analyses of patient-reported outcomes (PROs) may be performed on subpopulations as appropriate.

For each instrument, the compliance rate by visit will be calculated as the number of patients with completed assessments at each visit divided by the number of patients in ITT population at each visit. Compliance rates, reasons for noncompliance, and data collected for each instrument will be summarized by treatment arm.

6.9.1. Functional Assessment of Cancer Therapy Hepatobiliary Questionnaire (FACT-Hep)

FACT-Hep scores and their change from baseline will be summarized descriptively at each assessment time point. The change from baseline in FACT-Hep scores will be evaluated to investigate statistically significant differences between ramucirumab and pooled placebo and between merestinib and pooled placebo.

A mixed model repeated measure analysis will be applied to evaluate trajectories of pre-specified scores, domains, and items to investigate differences between ramucirumab and placebo and between merestinib and placebo; the model will include baseline scores as a covariate, time, treatment, and interaction terms of time by treatment and time by baseline scores as fixed effects. Time to deterioration (TtD) of selected scores, domains, and items will be characterized using the Kaplan-Meier method; between-treatment arm comparisons will be estimated using the Cox regression model (Cox 1972).

6.9.2. EuroQol 5-Dimension 5-Level (EQ-5D-5L)

For the EQ-5D-5L patient responses, based on their experiences during that day, the first 5 EQ-5D items form a 5-digit code uniquely identifying 1 of the 3125 possible health states. These unique 5-digit health states serve only as an index, or locator, into a pre-selected, country-specific value set to assign a utility value to this particular EQ-5D descriptive profile. These utility values are applied to the OS estimate for the estimation of quality-adjusted-life-years (QALYs) that are used to inform economic evaluations of health care interventions. No inference on these descriptive measures was intended by the questionnaire developers. Nonetheless, descriptive presentation of these data will be provided, by visit and by treatment, and will include numbers and proportions of responders on each of the 5 dimensions across each of the 5-level categorical responses (that is, mobility, self-care, usual activity, pain/discomfort, and anxiety/depression). The visual analogue scale (VAS) will be scored from 0 (worst imaginable health state) through 100 (best imaginable health state) to represent the patient's self-report concerning how bad or how good their health was during that day. For both the VAS scores and the produced utility values, descriptive statistics will include a measure of the central

tendency and a measure of dispersion (that is, these can be mean values and standard deviations [or standard errors, or 95% CIs] or if the data are skewed, median values and the 25th and 75th percentiles). Completion compliance of the EQ-5D-5L questionnaires will be described by assessment time point (including baseline, on-study, and short-term follow-up) by the number and percentage of patients who filled out a questionnaire (per patient, at least 1 question answered) over number of patients who are expected to complete the questionnaire at that time point. Reasons for noncompliance will be summarized.

6.10. Bioanalytical and Pharmacokinetic/Pharmacodynamic Methods

Planned pharmacokinetic and PK/pharmacodynamic (PD) analyses are specified in a separate standalone PK/PD analysis plan.

6.11. Safety Analyses

6.11.1. Extent of Exposure

The number of dose adjustments (dose omissions, reductions, delays, and increases), the number of cycles received, the duration of therapy, the cumulative dose, and dose intensity/relative dose intensity will be summarized by treatment arm. Dose intensity is defined as actual cumulative amount of drug taken / duration of treatment. Relative dose intensity is calculated as (actual amount of drug taken /amount of drug prescribed)*100%.

6.11.2. Adverse Events

All patients who receive at least one dose of any study treatment will be evaluated for safety and toxicity. Adverse event terms and severity grades will be assigned by the investigator using CTCAE v4.0X. In addition, AE verbatim text will also be mapped by the sponsor or designee to corresponding terminology within the Medical Dictionary for Regulatory Activities (MedDRA). Adverse events will be reported using a unified CTCAE/MedDRA reporting process:

- The CTCAE v4.0 term reported by the investigator will be mapped to the MedDRA preferred term (PT) and system organ class (SOC) of the corresponding MedDRA lowest level term (LLT), unless the reported CTCAE term is 'Other – specify'
- If the reported CTCAE term is 'Other – specify' the MedDRA LLT, PT and SOC mapped from the verbatim AE term will be used
- All listings and summaries will use the MedDRA LLT and the MedDRA PT resulting from this process.

A treatment-emergent adverse event (TEAE) is defined as an event that first occurred or worsened in CTCAE grade after the first dose of study treatment. The MedDRA LLT will be used in the treatment-emergent computation.

All observed AEs will be graded using CTCAE version 4.0. Adverse events, deaths and serious adverse events (SAEs) will be listed and summarized. Summaries for patients on therapy will include:

- Summary of all pre-existing conditions and AEs
- Summary of all TEAEs
- Summary of TEAEs possibly related to study treatment
- Summary of all TEAEs by maximum CTCAE grade
- Summary of TEAEs possibly related to study treatment by maximum CTCAE grade
- Summary of deaths on treatment or post-study treatment deaths possibly related to study treatment.

6.11.3. Deaths

All deaths recorded in this study will be included as part of the complete AE listing, where appropriate, and listed separately. A summary of deaths may be presented for all patients on therapy if there are a sufficient number of events for this to be deemed useful.

6.11.4. Clinical Laboratory Evaluation

Laboratory data (hematology, chemistry) will be listed for all patients on therapy. Abnormal results will be listed separately for all patients on therapy.

In addition to the investigator-reported AEs, all relevant hematology and chemistry laboratory values will be graded according to CTCAE version 4.0. These derived values will be included on the listings of laboratory data and summary tables will be produced in a similar manner to those created for the investigator reported AEs.

6.11.5. Vital Signs and Other Physical Findings

All vital signs including blood pressure, pulse, and temperature will be listed for all patients on therapy.

6.12. Subgroup Analyses

The primary analysis of PFS described in Section 6.8.1 will be conducted in each of the following subgroups:

- Primary Tumor Site (gall bladder versus intrahepatic cholangiocarcinoma versus extrahepatic cholangiocarcinoma or Ampulla of Vater)
- Geographic regions (North America or Europe versus Rest of World)
- Metastatic disease (yes versus no)
- MET mutation status
- MET immunohistochemistry
- Prior therapy (e.g. adjuvant, neoadjuvant, no prior therapy)
- Baseline ECOG performance status (0 verus 1)
- Cachexia (weight loss > 5% versus ≤ 5% over the prior 1 month; weight loss > 5% versus ≤ 5% over the prior 6 months)

6.13. Protocol Violations

Important protocol deviations that potentially compromise the data integrity and patients' safety will be listed. These deviations will include those defined by:

- Informed consent
- Inclusion/Exclusion Criteria
- Investigational Product
- Study Procedures
- Administrative/Oversight
- Safety
- Other

Based on the discussion with study team, the detailed description of each deviation within the above category and the method to identify each deviation will be listed in a separate document – Business Process Document: Important Protocol Deviations.

6.14. Interim Analyses and Data Monitoring

One interim analysis to assess safety will be conducted after a total of at least 75 evaluable patients have completed Cycle 1.

An interim efficacy analysis of OS will be conducted at the primary analysis of PFS, using gatekeeping approach.

6.15. Planned Exploratory Analyses

6.15.1. Biomarker Analysis

6.15.1.1. Analysis of Nonpharmacogenetic Biomarkers

Descriptive statistics will be provided for nonpharmacogenetic biomarkers collected from whole blood, serum and EDTA plasma samples for pre-treatment (baseline) and post-baseline time points when the samples were collected (See Appendix 4 of the clinical study protocol) by treatment arm (Arms A1, A2, B1 and B2). Descriptive statistics will include mean, standard deviation, median, minimum and maximum, or percentage if the biomarker is categorical.

Correlation between continuous biomarkers at baseline will be reported within each cohort (Arms A1, B1) and overall.

Scatter plot of survival time (both PFS and OS) against biomarker value at baseline will be provided by treatment arm in the translational research (TR) population. The TR population includes patients with non-missing values for at least one biomarker of interest at baseline .

Relationship between baseline biomarker measures and response to merestinib or ramucirumab will be investigated. Specifically, survival probability across time points and the median

survival time (95% CI) and survival time quartiles (95% CI) for both PFS and OS within each treatment arm will be estimated for patient subpopulations defined by biomarker value at baseline using the Kaplan-Meier method. In addition, Cox proportional hazards model will be employed to investigate whether the efficacy of either merestinib or ramucirumab treatment on PFS and OS is modified in patient subgroups defined by baseline biomarker values. If there is a sufficient number of patients in each stratum, then the HR between treatment arms for PFS and OS will be estimated using stratified Cox regression model (Cox 1972), stratified by the randomization stratification factors. Otherwise, the randomization stratification factors will be modeled as covariates in the Cox regression model to estimate the HR and 95% CI. The same stratification factors as described in Section **Error! Reference source not found.** will be adjusted for biomarker analysis.

The following patient subgroups will be considered:

- Patients with baseline biomarker value ≥ median vs. < median (for continuous biomarkers)
- Patients with zero baseline biomarker values vs. non-zero baseline biomarker values (for dichotomized biomarkers)

If other patient subgroups are considered, the change will be documented in the statistical analysis report.

Diagnostic plots and Subpopulation Treatment Effect Pattern Plots (STEPP) will be provided for continuous biomarkers to investigate an optimal cutpoint for the biomarkers.

6.15.1.2. Analysis of Pharmacogenetic Biomarkers

Pharmacogentic biomarkers may include DNA and RNA data. For DNA data, genome-wide association study (GWAS) or candidate gene study (CGS) may be considered aiming at identifying genetic markers with prognostic effect on survival time. Genotype by treatment interaction might be explored to investigate if any markers had predictive effect on merestinib or ramucriumab treatment as compared to placebo.

RNA-seq or microarray data, if available, will be analyzed to investigate the gene expression level change over time. Descriptive statistics for the expression level of genes of interest will be provided by treatment arm over time. Mixed effect Model Repeated Measurement analysis may be employed to investigate if treatment modifies the change of gene expression level over time. Association of survival time with baseline gene expression will be explored to identify gene expression profiles with prognostic or predictive effect.

6.15.2. Relationship of Pretreatment Weight Loss with Efficacy Variables

Relationship of pretreatment weight loss (Section 6.12) with primary and secondary treatment response outcomes will be evaluated. The treatment response outcomes include PFS, OS, DoR, and ORR. All time-to-event variables will be analyzed by Cox regression models; ORR will be compared between pretreatment weight loss (weight loss > 5% versus ≤ 5%) using Fisher's exact

test. The Cox regression model will include pretreatment weight loss, treatment (experimental arms and pooled control), age, gender, primary tumor site, prior therapy, and baseline ECOG performance status.

6.16. Clinical Trial Registry Analyses

Additional analyses will be performed for the purpose of fulfilling the Clinical Trial Registry (CTR) requirements.

Analyses provided for the CTR requirements include the following:

- Summary of adverse events, provided as a dataset which will be converted to an XML file. Both Serious Adverse Events and 'Other' Adverse Events are summarized: by treatment group, by MedDRA preferred term.
- An adverse event is considered 'Serious' whether or not it is a treatment emergent adverse event (TEAE).
- An adverse event is considered in the 'Other' category if it is both a TEAE and is not serious. For each Serious AE and 'Other' AE, for each term and treatment group, the following are provided:
    - the number of participants at risk of an event
    - the number of participants who experienced each event term
    - the number of events experienced.
- Consistent with www.ClinicalTrials.gov requirements, 'Other' AEs that occur in fewer than 5% of patients/subjects in every treatment group may not be included if a 5% threshold is chosen (5% is the minimum threshold).
- AE reporting is consistent with other document disclosures for example, the CSR, manuscripts, and so forth.

In addition, the following rules apply in order to meet the requirement for participant flow and accurately represent study completion.

Study discontinuation reason	Completed	Not Completed
Death due to any cause	X	
Patient alive and on study at conclusion, but off treatment	X	
Lost to follow-up		X
Withdrew consent **TO STUDY PARTICIPATION** (patient or physician)		X
On study treatment at study conclusion		X

7. Unblinding Plan

The purpose of this unblinding plan is to maintain the scientific integrity of the study. Access to study data will be strictly controlled until the final database lock occurs.

Randomization will occur at each patient's baseline visit using an interactive web response system (IWRS). Assignment to treatment groups will be determined by stratified randomization. Only a minimum number of Lilly personnel with the IWRS group will have access to the randomization algorithm and treatment assignments before the study is complete.

To preserve blinding of the study, a minimum number of sponsor personnel will see treatment assignments before final OS analyses are completed.

Emergency unblinding for AEs may be performed through an IWRS, which may supplement or take the place of emergency codes generated by a computer drug-labeling system. This option may be used only if the patient's well-being requires knowledge of the patient's treatment assignment. All calls resulting in an unblinding event are recorded and reported by the IWRS.

The investigator should make every effort to contact the Lilly CRP prior to unblinding a patient's treatment assignment. If a patient's treatment assignment is unblinded, Lilly must be notified immediately by telephone.

Every effort will be made to blind both the patient and the investigator to the treatment assignment, but the inadvertent unblinding of a patient may occur. A blinded study design is known to be imperfect because the potential for individual unblinding exists due to treatment-related signs and symptoms. If a patient or investigator is unblinded, the unblinding alone will not be sufficient cause for that patient to be discontinued from study treatment or excluded from any safety or efficacy analyses.

Additionally, there may be ethical reasons to have the patient remain on study treatment. For patients to continue on study treatment in the event of unblinding, the investigator must obtain specific approval from a Lilly CRP for the patient to continue in the study.

Patients and investigators will remain blinded to study treatment with no sharing of efficacy information among sites until final OS analyses are completed. Treatment assignment will be blinded in the reporting database except for a minimum number of sponsor personnel required to perform the interim safety, primary PFS analysis, and interim efficacy analysis of OS. This will ensure unblinded aggregate efficacy results are not available outside of the Assessment Committee until the time of final OS analysis. Patients and investigators may be unblinded after final OS analyses are completed.

Only the assessment committee (AC) is authorized to evaluate unblinded interim safety analysis, primary PFS analysis, and interim efficacy analysis of OS. All study team members will remain blinded to interim safety, primary PFS results, and interim OS results, except in the case that AC concludes that significant efficacy has been demonstrated at the primary PFS analysis and/or interim efficacy analysis of OS, a limited number of people will need to have access to the unblinded data to perform analyses and make critical decisions (for example, regulatory

submission). This group of people will be identified and documented prior to the database lock for the primary PFS analysis. No by-patient level treatment data will be accessible to anyone else (for example, the rest of study team and investigators) until the database lock for the final analysis of OS. Study sites will receive information about interim results ONLY if they need to know for the safety of their patients.

The limited number of people (outside of AC) who will need to have access to the unblinded data to perform analyses and make critical decisions are:

- Study CRP/CRS
- Regulatory scientist
- PK scientist
- Health outcome scientist
- Biomarker scientist
- Product team lead

8. References

Cox DR. Regression models and life-tables. *J Royal Stat Soc Ser B*. 1972;34(2):187-220.

Leo Document ID = f2b06658-281e-489d-b60f-cd39feab217b

Approver: PPD
Approval Date & Time: 19-Mar-2018 20:09:04 GMT
Signature meaning: Approved