Intercept

Statistical Analysis Plan

OBETICHOLIC ACID (OCA)

747-302

A Phase 4, Double-Blind, Randomized, Placebo-Controlled, Multicenter Study Evaluating the Effect of Obeticholic Acid on Clinical Outcomes in Subjects with Primary Biliary Cholangitis

THE COBALT STUDY

Clinical Outcomes with <u>OB</u>eticholic <u>A</u>cid in <u>L</u>iver <u>T</u>reatment (COBALT)

EudraCT Number: 2014-005012-42

ClinicalTrials.gov Identifier: NCT02308111

<u>Sponsor</u>

Intercept Pharmaceuticals, Inc.
305 Madison Avenue
Morristown, NJ 07960

USA

16.1.9   DOCUMENTATION OF STATISTICAL METHODS

747-302 Statistical Analysis Plan v5 30Mar2022

747-302 Clinical Pharmacology Analysis Plan v1 12Apr2022

747-302 Statistical Analysis Plan for Interim Analysis v1 20Aug2020

OBETICHOLIC ACID (OCA)

747-302

THE COBALT STUDY

Clinical Outcomes with OBeticholic Acid in Liver Treatment (COBALT)

A Phase 4, Double-Blind, Randomized, Placebo-Controlled, Multicenter Study Evaluating the Effect of Obeticholic Acid on Clinical Outcomes in Subjects with Primary Biliary Cholangitis

Statistical Analysis Plan

Author:	PPD Ph.D. Associate Director, Biostatistics, Intercept Pharmaceutical
Protocol Version and Date:	Version 6: 05 November 2020
Analysis Plan Version:	5
Analysis Plan Date:	30 March 2022

APPROVAL

Upon review of this document, including table and listing shells, the undersigned approves the Statistical Analysis Plan. The analysis methods and data presentation are acceptable.

PPD PhD
Associate Director, Biostatistics
Intercept Pharmaceuticals, Inc.

PPD

Signature   4/5/2022
                                                 Date

PPD PhD
PPD Biostatistics and Data
Management
Intercept Pharmaceuticals, Inc.

PPD

Signature   4/4/2022
                                                 Date

PPD MD
Senior Director, Scientific Affairs
Intercept Pharmaceuticals, Inc.

PPD

Signature   4/4/2022
                                                 Date

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

Abbreviation or Specialist Term	Explanation
AE	adverse event
AESI	adverse event of special interest
ALP	alkaline phosphatase
ALT	alanine aminotransferase
AST	aspartate aminotransferase
ATC	Anatomic Therapeutic Chemical
CI	confidence interval
CIF	cumulative incidence function
CK-18	cytokeratin-18
CLIF-C ADs	Chronic Liver Failure Consortium Acute Decompensation score
CP	Child-Pugh
CRF	case report form
CRP	C-reactive protein
CS	clinically significant
CSPH	clinically significant portal hypertension
CSR	clinical study report
DMC	Data Monitoring Committee
CMH	Cochran–Mantel–Haenszel
ECG	electrocardiogram
EDC	electronic data capture
EHR	electronic health records
ELF	enhanced liver fibrosis
EMA	European Medicines Agency
FGF-19	fibroblast growth factor-19
GGT	gamma-glutamyl transferase
HCC	hepatocellular carcinoma
HR	hazard ratio
ICH	International Conference on Harmonisation
INR	international normalized ratio
IP	investigational product
IPCW	Inverse Probability of Censoring Weighting
IPTW	Inverse Probability of Treatment Weights

Abbreviation or Specialist Term	Explanation
IWRS	interactive web response system
KM	Kaplan-Meier
LS	least-squares
MELD	Model for End-Stage Liver Disease
MMRM	Mixed-effect Model Repeated Measures
MRS	Mayo risk score
NCS	not clinically significant
OCA	obeticholic acid
ORAE	on-risk adverse event
PBC	primary biliary cholangitis
PD	pharmacodynamic
PK	pharmacokinetic
PMN	polymorph leucocytes
PT	preferred term
QTcF	QT interval corrected by the Fridericia's formula
SAE	serious adverse event
SAP	statistical analysis plan
SD	standard deviation
SEM	standard error of the mean
SOC	system organ class
TE	transient elastography
TEAE	treatment-emergent adverse event
TNF-α	tumor necrosis factor-alpha
UDCA	ursodeoxycholic acid
UK	United Kingdom
ULN	upper limit of normal
US	United States (of America)
USPI	United States Prescribing Information
VLDL	very low density lipoprotein

1.   SCOPE

This statistical analysis plan (SAP) describes the statistical analyses and data presentations planned for Protocol 747-302, Clinical Outcomes with OBeticholic Acid in Liver Treatment (COBALT), Version 6. It provides a detailed description of the strategy, rationale, and statistical methods to be used to meet the study objectives, as well as additional details to the statistical analyses that were described in the study protocol. Any deviations from the methods specified in this SAP will be documented in the clinical study report (CSR). If after the database lock additional analyses are required to supplement the planned analyses described in this SAP or any addendums, they will be identified as post-hoc in the CSR.

2.   STUDY OBJECTIVES AND HYPOTHESES

2.1.   Study Objectives

To support the primary objective, the primary efficacy endpoint has been expanded beyond what was outlined in Protocol Version 6. Further details are presented in Section 14.1.

2.1.1.   Primary Objective

The primary objective of this study is to compare the effect of obeticholic acid (OCA) to placebo, in conjunction with established local standard of care, on clinical outcomes in subjects with primary biliary cholangitis (PBC) as measured by time to first occurrence of the following adjudicated events, derived as a composite event endpoint (including all Group 1 – Group 3 events).

Group 1:

- Death (all-cause)

- Liver transplant

- Hospitalization (as defined by a stay of 24 hours or greater) for new onset or recurrence of:

  - Variceal bleed

  - Hepatic encephalopathy (as defined by a West Haven score of ≥ 2)

  - Spontaneous bacterial peritonitis (confirmed by diagnostic paracentesis OR presence of $>250/mm^3$ polymorph leucocytes [PMNs] in the ascitic fluid)

  - Bacterial empyema (confirmed by diagnostic thoracentesis OR presence of $>250/mm^3$ PMNs in the pleural fluid)

- Uncontrolled or refractory ascites (requiring large volume paracentesis)

- Portal hypertension syndromes:

  - Hepatorenal syndrome (as defined by International Ascites Club [Angeli 2019])

  - Portopulmonary syndrome

  - Hepatopulmonary syndrome

- Model of end-stage liver disease (MELD)-Na score ≥15 if MELD-Na <12 at baseline

- MELD score ≥15 if MELD-Na ≥12 at baseline

Group 2:

- Progression to decompensated liver disease (for subjects without decompensation at baseline), prioritized as follows:

  - New onset of:

    - Hepatic hydrothorax

    - Variceal bleeding

    - Ascites requiring treatment with sodium restriction, diet modification, or diuretics

  - Hepatic encephalopathy requiring lactulose and/or rifaximin

  - New onset of:

    - Child-Pugh ≥7

    - Total bilirubin >3 mg/dL

Group 3:

- Progression to clinical evidence of portal hypertension without decompensation (for subjects without decompensation or clinical evidence of portal hypertension at baseline)

  - Endoscopic evidence of portal hypertension without bleeding

    - Gastroesophageal varices (requiring banding or progression to large varices if no or small varices were observed at baseline)

    - Portal hypertensive gastropathy

  - Platelets $<150 \times 10^9$/L with splenomegaly and/or with transient elastography >15 kPa

2.1.2. Secondary Objectives

- To assess the effect of OCA compared to placebo on time to first occurrence of death, liver transplant, MELD-Na score ≥15 if MELD-Na< 12 at baseline, MELD score ≥15 if MELD-Na ≥12 at baseline, uncontrolled or refractory ascites, portal hypertension syndromes (hepatorenal syndrome, portopulmonary syndrome, hepatopulmonary syndrome), or hospitalization for new onset or recurrence of variceal bleed, hepatic encephalopathy, spontaneous bacterial peritonitis, or bacterial empyema.

- To assess the effect of OCA compared to placebo on time to first occurrence of death, liver transplant, MELD ≥15, uncontrolled ascites, or hospitalization for new onset or recurrence of variceal bleed, hepatic encephalopathy, or spontaneous bacterial peritonitis.

- To assess the effect of OCA compared to placebo on time to first occurrence of each individual component of the primary endpoint as listed above.

- To assess the effect of OCA compared to placebo on time to occurrence of liver-related death.

- To assess the effect of OCA compared to placebo on progression to cirrhosis.

- To assess the effect of OCA compared to placebo on time to occurrence of hepatocellular carcinoma (HCC).

- To assess the effect of OCA compared to placebo on disease progression via the following:

  – Liver biochemistry

  – Markers of inflammation and fibrosis

- To characterize the PK of OCA and its conjugates in a subset of subjects.

- To assess health outcomes and pharmacoeconomics including cost-effectiveness, resource utilization, and quality of life measures in subjects treated with OCA compared to placebo.

- To assess the safety and tolerability in subjects treated with OCA compared to placebo.

2.2. Hypotheses

2.2.1. Primary Hypothesis

The study is intended to assess whether there is a difference between OCA and placebo in the time from baseline to the first occurrence of the composite event endpoint (as defined in Section 2.1.1, Groups 1 – 3).

Two-sided hypotheses are expressed in terms of:

- The null hypotheses (H_0) that the two hazard rates (h_{OCA}, $h_{Placebo}$) are the same, that is, that there is no difference between the OCA and placebo populations in the probability of an event at any timepoint.

- The alternative hypothesis (H_1) that the two hazard rates are not the same ($h_{OCA} \neq h_{Placebo}$)

2.2.2. Secondary Hypothesis

The secondary hypotheses are intended to assess whether there is a difference between OCA and placebo in the time from baseline to the first occurrence of the composite event endpoint (as defined in Section 2.1.1) and other time-to-event endpoints.

Two-sided hypotheses are expressed in terms of:

- The null hypotheses (H$_0$) that the two hazard rates (h$_{OCA}$, h$_{Placebo}$) are the same, that is, that there is no difference between the OCA and placebo populations in the probability of an event at any timepoint.

- The alternative hypothesis (H$_1$) that the two hazard rates are not the same (h$_{OCA} \neq$ h$_{Placebo}$)

3. SUMMARY OF STUDY DESIGN

This Phase 4, double-blind, randomized, placebo-control, multicenter study will evaluate the effect of OCA on clinical outcomes in subjects with PBC who are at higher risk of liver-related clinical complications.

Eligible subjects will have a diagnosis of PBC with bilirubin levels >ULN and ≤5x ULN and/ or ALP >3× ULN. Subjects will be screened twice during a 1- to 8-week Screening period prior to entering the study to allow for the collection of repeat serum chemistry samples (at least 2 weeks apart) to confirm pretreatment alkaline phosphatase (ALP) and total bilirubin values.

Investigational product (IP) will be taken orally, once daily for the majority of subjects; dose and frequency will be modified for subjects with cirrhosis (including subjects progressing to cirrhosis during the study) and classified as Child-Pugh B or C. The randomization will be stratified by ursodeoxycholic acid (UDCA) treatment (yes/ no) and baseline bilirubin categories (> upper limit of normal [ULN]/ ≤ULN). A minimum of 30% of subjects will have elevated bilirubin (>ULN) at Screening.

The study is event driven and will continue until approximately 127 adjudicated primary endpoint events have been accrued in unique subjects, or until the Sponsor (eg, based on a recommendation from the Data Monitoring Committee (DMC)) terminates the study. Subjects are expected to be followed for a minimum of approximately 6 years. Figure 1 below displays the study design diagram. Schedules of study procedures are outlined in Table 1 (Screening to Month 12) and Table 2 (Year 2 to End-of-Study Endpoints) of the study protocol, where more details are provided.

Figure 1:   Schematic Diagram Study 747-302
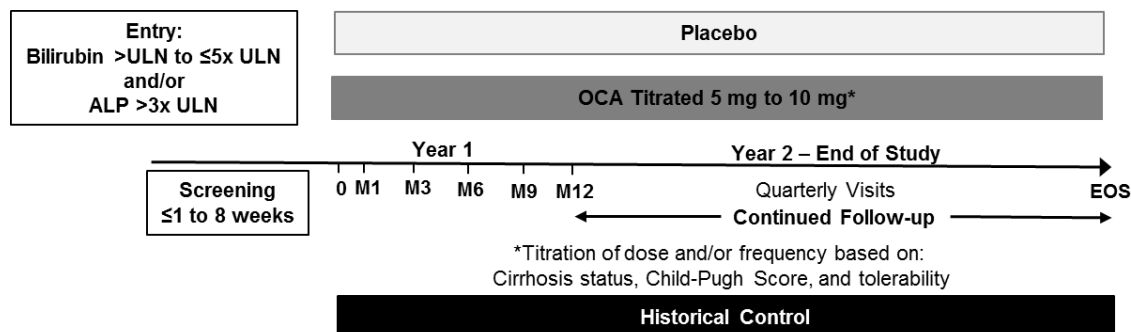
EOS = end of study; M = month, OCA = obeticholic acid; ULN = upper limit of normal
Initial dose titration of investigational product should occur at the Month 3 visit, or any study visit thereafter for subjects on all dosing regimens, based on tolerability and biochemical response (Up-titration should be considered when ALP and/ or total bilirubin are >ULN). Subsequent dose titration(s) for subjects classified as Child Pugh B or Child-Pugh C and following a modified dosing schedule may occur no earlier than 6 weeks after the previous dose titration.

Dosing frequency will be determined by the presence or absence of cirrhosis and, if cirrhosis is present, by Child-Pugh Score as described below:

	Planned Dosing Regimen	
	Standard	**Modified**
	Noncirrhotic/ Child-Pugh A	**Child-Pugh B or Child-Pugh C**
Starting Dose [a] (Day 0)	5 mg daily	5 mg once weekly
Titration 1 [b] (≥Month 3)	10 mg daily	5 mg twice weekly
Titration 2 [b] (≥6 weeks after Titration 1)	NA	10 mg twice weekly

[a] Starting dose based on subject's cirrhosis status and Child-Pugh Score at Screening.
[b] Planned titration regimen is shown; however, the titration of dose and/or frequency is dependent on subject tolerability, biochemical response, and/or changes in cirrhosis status or Child-Pugh Score at any time during the study
[c] Dosing per the twice weekly schedule must be at least 3 days apart.

With the exception of liver transplant, if a subject experiences a suspected or confirmed clinical outcomes event, the subject should continue the regular visit schedule and continue taking IP as long as the subject has not met any of the criteria for discontinuation outlined in the protocol. Subjects who discontinue IP are expected to be followed through to study closure (or at the discretion of the Sponsor).

3.1.   Randomization and Blinding

This study will be conducted in a double-blind, placebo-controlled manner. Allocation to one of two treatment groups (OCA or placebo) will occur on a 1:1 ratio across sites and will be stratified by UDCA treatment (yes/no) and mean baseline bilirubin categories (>ULN/≤ULN), as specified by the central laboratory. The randomization will be based on a predefined randomization code (generated by the Sponsor or designee) using an interactive web response system (IWRS). The IWRS will also serve as an investigational product inventory and management system and may be integrated with the study database.

The unblinded data access during DMC and interim analysis is described in Safety Review Committee Charter.

3.2.   Sample Size Determination

The following assumptions were used in the Protocol Version 6 sample size calculations:

- Exponential survival curves, placebo survival estimate of 0.6 at 8 years with a hazard ratio of 0.60, and total study duration of 10 years (from first subject enrolled), allowing for 4 years of subject accrual and 6 years of follow up.
- Subjects will be randomized in a 1:1 ratio to placebo or OCA.
- The 2 treatment groups will be compared using a 2-sided log rank test at the 5% level of significance.
- Two interim analyses and one final analysis are planned, with interim analyses occurring after the accrual of 50% and 75% of clinical outcome events, respectively.
- A dropout rate of 10% is assumed

Based on the randomization ratio, significance level, and assumed hazard ratio, a total of 127 events (both groups combined) was to provide 80% power to demonstrate a statistically significant difference between OCA and placebo on time to liver-related outcomes, including all-cause mortality.

In addition, based on the remaining assumptions stated above, it was estimated that approximately 428 subjects would be enrolled to attain 127 events.

Sample size calculations were based on the following composite endpoint: time to first occurrence of death, liver transplant, MELD ≥15, uncontrolled ascites, or hospitalization for new onset or recurrence of variceal bleed, hepatic encephalopathy, or spontaneous bacterial peritonitis.

Due to enrollment challenges and difficulties to keep patients in the study, the trial will be terminated early by sponsor. To better evaluate the benefit-risk ratio, the composite endpoint will be revised to include additional components that occur earlier in the progression of PBC

4.   ANALYSIS POPULATIONS AND APPROACHES TO ANALYSIS

4.1.   Intent-to-Treat (ITT) Population

The ITT Population includes all randomized subjects who receive any amount of investigational product (OCA or placebo). The ITT Population will be the primary population used for efficacy analyses. Treatment assignment will be based on the randomized treatment.

4.2.   Per Protocol (PP) Population

The PP Population includes ITT subjects who do not have any major protocol deviations that potentially affect the primary efficacy conclusion. Major protocol deviations that exclude subjects from the PP population are:

- Subject randomized into the study despite not meeting inclusion criteria or meeting exclusion criteria.

- Potential overdose, i.e. investigational product compliance >120%, unless there's clear documentation.

- Subject met protocol criteria for down-titration, but dose was not adjusted.

- Prohibited change in concomitant medication whilst actively on the study, including commercial OCA.

- SAE not reported or not reported in required timeframe.

- Failure to submit local subject level safety events per IRB/EC policy.

- Withdraw criteria met, but subject not discontinued.

The PP Population will be used for the sensitivity analyses of the primary and key secondary efficacy endpoints. Treatment assignment will be based on the randomized treatment. PP patients list will be determined prior to unblinding.

4.3.   Safety Population

The Safety Population includes all randomized subjects who receive any amount of investigational product (OCA or placebo). The Safety Population will be the primary population used for safety analyses and treatment assignment will be based on the treatment actually received.

4.4.   Pharmacokinetic (PK) Population

The PK Population includes all OCA subjects who have at least one confirmed fasted analyzable sample. Subjects must have been fasting for approximately 8 hours prior to the visit and must not have any major protocol deviations that potentially affect exposure levels. The PK Population will be used for OCA PK analyses.

5.   SUBJECT DISPOSITION

Subject disposition will be tabulated by treatment group and overall and will include the number and proportion of subjects screened and enrolled, as well as subject discontinuation status and reasons for discontinuing treatment or study. Discontinuation status will be additionally presented for subjects with and without a primary endpoint event. Subject disposition will be summarized for all analysis populations defined in Section 4.

Subjects are considered to have discontinued from the study if they discontinued study visits and did not consent to follow-up contact or medical record review, or if they withdrew consent for follow-up contact or medical record review. Subjects are considered to have started commercial Ocaliva if they discontinued treatment or study visits due to initiating commercial Ocaliva, or if Ocaliva is recorded as a concomitant medication.

5.1.   Protocol Deviations

The Investigator is not permitted to deviate from the protocol in any significant way without proper notification to the Sponsor (or designee). Only the Sponsor may amend the protocol. Any change in study conduct considered necessary by the Investigator will be made only after consultation with the Sponsor, who will then issue a formal protocol amendment to implement the change and obtain regulatory approval. The only exception is when the Investigator considers a subject's safety to be compromised if immediate action is not taken.

Protocol deviations identified during the conduct of the study will be summarized and listed.

6.   DEMOGRAPHICS AND BASELINE DISEASE CHARACTERISTICS

Demographic variables (eg, age at informed consent, age categorized as < 65 years and $\geq$ 65 years and by decades, sex, race, ethnicity, and geographic region), other baseline characteristics (eg, weight, height, body mass index, BMI categorized as < 30 kg/m^2 and $\geq$ 30 kg/m^2, use of UDCA at screening (Yes/No)), and key liver function test results at baseline (eg, ALP, total

bilirubin, albumin, platelets, international normalized ratio (INR), GGT, creatinine) will be summarized and presented by treatment and overall for each analysis population.

Summaries of PBC disease characteristics will use data collected from the PBC Disease History electronic case report form (eCRF). Variables from the PBC Disease History eCRF include the following:

- Age at PBC diagnosis

- Duration of PBC in years at time of informed consent

- PBC stage of most recent liver biopsy

- History of PBC related pruritus (Yes/No)

- Severity of most recent pruritus event

    - Ongoing at screening

- History of PBC-related fatigue (Yes/No)

    - Ongoing at screening

- Overall severity of PBC-related fatigue

- Use of UDCA (Never/Previous/Current)

- Duration of previous/current UDCA use in years at time of informed consent

For categorical variables, the number and percentage of subjects within each category, including a category for missing data, will be presented. For continuous variables, summary statistics will include the number of subjects and the mean, standard deviation (SD), standard error of the mean (SEM), median, 25th and 75th quartiles, minimum, and maximum values.

No inferential statistical comparisons will be performed.

7.   PRIOR THERAPY AND MEDICAL HISTORY

Prior therapy and medical history includes any therapy or diseases that occurred or any medication taken prior to baseline date.

Prior therapy will be mapped to Anatomic Therapeutic Chemical (ATC) class and preferred term (PT) using WHO Drug Dictionary version Global B3, March 2020 (or later version) and summarized by ATC class, preferred term, and treatment in the Safety Population. Summaries will be ordered by descending order of incidence of ATC class and preferred term within each ATC class.

Medical history will be mapped to PTs and system organ classes (SOCs) using MedDRA dictionary version 23.1 (or later version) and summarized by SOC, PT, and treatment group in the Safety Population. Summaries will be ordered by descending order of incidence of SOC and preferred term within each SOC.

8.   EFFICACY ANALYSES

The efficacy endpoints and the planned analyses of these endpoints are described below.

8.1.   Primary Efficacy Endpoint

The following table identifies the elements of the intercurrent events and strategies to address for primary efficacy endpoint.

Primary Estimand	
Population	Adult subjects with PBC who are at higher risk of liver-related clinical complications.

Endpoint (variable)	Compare the effect of OCA to placebo, in conjunction with established local standard of care, on the composite clinical outcomes as measured by time from randomization to the first occurrence of the following post-randomization events:
	
	Group 1 (applies to all subjects):
	
	• Death (all-cause)
	• Liver transplant
	• Hospitalization (as defined by a stay of 24 hours or greater) for new onset or recurrence of:
	  – Variceal bleed
	  – Hepatic encephalopathy (as defined by a West Haven score of ≥ 2)
	  – Spontaneous bacterial peritonitis (confirmed by diagnostic paracentesis OR presence of $>250/mm^3$ PMNs in the ascitic fluid)
	  – Bacterial empyema (confirmed by diagnostic thoracentesis OR presence of $>250/mm^3$ PMNs in the pleural fluid)
	• Uncontrolled or refractory ascites (requiring large volume paracentesis)
	
	• Portal hypertension syndromes (hepatorenal, portopulmonary, and hepatopulmonary)
	
	• MELD-Na score ≥ 15 (for subjects with baseline MELD-Na score <12)
	• MELD score ≥ 15 (for subjects with baseline MELD-Na score ≥ 12)
	
	Group 2 (applies to subjects without decompensation at baseline):
	• Progression to decompensated liver disease (for subjects without decompensation at baseline). See Section 2.1.1 Group 2 and Section 8.1.1 for prioritization list and analysis.
	
	Group 3 (applies to subjects without decompensation or clinical evidence of portal hypertension at baseline):
	• Progression to clinical evidence of portal hypertension without decompensation (for subjects without decompensation or clinical evidence of portal hypertension at baseline). See Section 2.1.1 Group 3 and Section 8.1.1 for prioritization list and analysis.

Primary Estimand		
Intercurrent events/ Strategies to address	Use of commercial OCA as concomitant medication or IP/study visit discontinuation due to commercial OCA use	Treatment policy will use for primary analysis: Treatment effect regardless of intercurrent event.
	IP discontinuation due to other reasons prior to termination of the study and continue to follow the regular visit schedule through to study closure	Treatment policy will use for primary analysis: Treatment effect regardless of intercurrent event.
Population-level summary	Log rank test of clinical outcomes of OCA group vs. placebo group.	

If at a given timepoint some components of the MELD or MELD-Na scores are available but others are missing, the missing components will be imputed as the most recent measure within a year prior to the timepoint (LOCF). If no measure is available within a year prior to the timepoint, the missing component will be imputed as in Table 1.

Table 1: **MELD and MELD-Na Score Imputations for Missing Components at Any Given Timepoint**

Measure	Imputed Value
Total bilirubin (mg/dL)	LLN
International Normalized Ratio (INR)	1
Serum creatinine (mg/dL)	LLN
Serum sodium (mmol/L)	ULN
Serum glucose (mg/dL)	LLN

8.1.1.   Primary Analyses

The primary efficacy analyses will compare OCA to placebo with respect to the primary efficacy endpoint using the ITT Population. Only adjudicated events will be included in the analysis. The number and percent of subjects censored and with events will be presented. Descriptive statistics will be presented for the time-to-event endpoints.

If a subject experiences more than one event in the composite primary endpoint (all Group 1 – Group 3 events), only the event that occurs first will be included in the analysis. Subjects with no data after randomization will be censored on Day 1 (first day of IP dosing). Subjects who do not experience an event will be censored at the time of their last contact. Last contact is the date of discontinuation from regularly-scheduled study visits for subjects who did not consent to follow-up and the date of discontinuation from contact visits (medical record review / semi-annual phone calls) for patients who consented to follow-up. The total number of patient years observed, patient years and percentage of patient time accrued after discontinuation of IP, and

patient years and percentage of patient time accrued after initiation of commercial OCALIVA will be presented overall and by treatment group.

The 2 treatment groups will be compared using the log rank test stratified by the randomization stratification factors, conducted at the 2-sided alpha level of significance described in Section 8.6. Kaplan-Meier (KM) estimates of the distribution of the time-to-event will be tabulated and graphed by treatment group. The tabulation will include the KM estimate of the 25th, 50th (median), and 75th percentiles and corresponding 2-sided 95% CIs, where the percentiles can be estimated. KM tabulations will present time to event in days; KM graphs will present time to event in months. The hazard ratio and 95% CI will be determined based on a Cox regression model stratified by randomization strata to estimate the magnitude of the effect. The proportionality of hazards will be assessed using Schoenfeld residuals.

8.1.2.   Sensitivity Analyses

The following sensitivity analyses will be carried out on the primary endpoint, with different population, strategy to handle the intercurrent events, or summary.

No.	Population	Strategies to address intercurrent events	Population-level summary
1	Subset of primary population, who had no major protocol violations.	Same as primary analysis	Same as primary analysis
2	Same as primary analysis	Same as primary analysis	The 2 treatment groups will be compared using the win ratio (Pocock 2012) using the unmatched method within the bilirubin randomization strata. The 95% confidence interval will be calculated using a bootstrapping method and a p-value will be calculated using the stratified Finkelstein-Schoenfeld test (Finkelstein 1999).
3	Same as primary analysis	Data collected after the intercurrent events to be excluded (while on treatment policy)	For subjects who didn't experience an event before the intercurrent event, multiple imputation will be carried out for time to event based on the available data after the intercurrent event. Details is included in the appendix.
4	Same as primary analysis	Exclude data after 90 days after a patient discontinues IP.	Same as primary analysis.
5	Same as primary analysis	Include only data prior to when a patient starts commercial OCALIVA.	Same as primary analysis.
6	Same as primary analysis	Include only data prior to when OCA was commercially available in the subject's country.	Same as primary analysis.
7	Same as primary analysis	Hypothetical strategy, ie, as if the intercurrent events didn't happen.	Hazard ratio and 95% CI will be estimated using a weighted Cox regression model with Sandwich estimator. The derivation of the weights based on Inverse Probability of Censoring Weighting (IPCW) is detailed in the appendix.

8.1.2.1.   Win Ratio

The win ratio analysis is intended to assess whether there is a difference between OCA and placebo in the hierarchical event endpoint (as defined in Section 2.1.1) while considering the clinical importance order and relative timing of its components.

In the win ratio analysis, two-sided hypotheses are expressed in terms of:

- The null hypotheses (H0) that the win ratio $R_w = 1$.

- The alternative hypothesis (H1) that the win ratio $R_w \neq 1$.

The 2 treatment groups will be compared using the win ratio (Pocock 2012) using the unmatched method within the bilirubin randomization strata. The win ratio represents the odds of having a more favorable outcome versus a less favorable outcome when assigned to the OCA treatment group compared to the placebo treatment group. If the estimated win ratio is greater than 1 then treatment with OCA is considered to be more favorable than treatment with placebo.

8.1.2.1.1.   Calculation of the Win Ratio

The components of the composite endpoint are prioritized as follows:

Group 1:

1	• Death (all-cause)
2	• Liver transplant
3	• Hospitalization (as defined by a stay of 24 hours or greater) for new onset or recurrence of: – Variceal bleed – Hepatic encephalopathy (as defined by a West Haven score of ≥2) – Spontaneous bacterial peritonitis (confirmed by diagnostic paracentesis OR presence of >250/mm^3 PMNs in the ascitic fluid) – Bacterial empyema (confirmed by diagnostic thoracentesis OR presence of >250/mm^3 PMNs in the pleural fluid)
4	• Uncontrolled or refractory ascites (requiring large volume paracentesis)
5	• Portal hypertension syndromes: – Hepatorenal syndrome – Portopulmonary syndrome – Hepatopulmonary syndrome
6.1	• MELD-Na score ≥15 when both subjects have a baseline MELD-Na score <12
6.2	• MELD score ≥15 when one or more subjects have a baseline MELD-Na score ≥12

Within each baseline bilirubin strata, $k \in (1, 2)$, each subject in the treatment group will be compared to each subject in the placebo group. For each pair, the time to the most important

event (eg, all-cause death) determines the "winner." If this comparison is non-informative, then the time to the second-most important event (eg, liver transplant) is compared to determine a "winner," and so on.

For each composite endpoint c, each subject i has (T_{ic}, C_{ic}), a time of the event and time of censoring (only one observed).

When comparing treated subject t to placebo subject p on time-to-event c,

- The treated subject "wins" if
    - (T_{tc}, C_{tc}) > T_{pc}
- The treated subject "loses" if
    - T_{tc} < (T_{pc}, C_{pc})
- The comparison is considered non-informative if
    - neither subject has event c,
    - min(C_{tc}, C_{pc}) < min(T_{tc}, T_{pc}), or
    - $T_{tc} = T_{pc}$.

For pairs in which all priority-ranked endpoints are non-informative, the win/loss will be determined as follows:

1. Progression to decompensated liver disease if neither subject has decompensation at baseline. (Group 2)

2. If tiebreaker #1 is non-informative or not applicable: Progression to clinical evidence of portal hypertension without decompensation if neither patient has decompensation or clinical evidence of portal hypertension at baseline. (Group 3)

Tiebreaker #1: progression to decompensated liver disease (for pairs in which both subjects do not have decompensation at baseline):

The components of progression to decompensated liver disease are prioritized as follows (Group 2):

1	• New onset of:
	    – Hepatic hydrothorax
	    – Variceal bleeding
	    – Ascites requiring treatment with sodium restriction, diet modification, or diuretics

2	• Hepatic encephalopathy requiring lactulose/or rifaximin
3	• New onset of Child-Pugh ≥7 • New onset of total bilirubin >3 mg/dL

For each pair, the time to the subjects' first event from the first priority category determines the "winner." If this comparison is non-informative, then the time to first occurrence of Child-Pugh ≥7 and/or total bilirubin >3 mg/dL is compared to determine a "winner".

Tiebreaker #2: progression to clinical evidence of portal hypertension without decompensation (for pairs in which both subjects do not have decompensation or clinical evidence of portal hypertension at baseline):

The components of progression to clinical evidence of portal hypertension are prioritized as follows (Group 3):

1	• Endoscopic evidence of portal hypertension without bleeding
	   – Gastroesophageal varices (requiring banding or progression to large varices if no or small varices were observed at baseline)
	   – Portal hypertensive gastropathy

2	• Platelets <150 x 10^9/L with splenomegaly and/or with transient elastography >15 kPa

Once all pairs of OCA and placebo-treated subjects are classified as a win or loss for the treatment group (or as non-informative, if none of the tiebreakers determine a winner), the win ratio will be determined.

The win ratio is the total number of wins for the OCA-treated group divided by the total number of losses for the OCA-treated group: $R_w = {N_w}/{N_l}$. The 95% confidence interval will be calculated using a bootstrapping method and a p-value will be calculated using the stratified Finkelstein-Schoenfeld test (Finkelstein 1999).

The number and percentage of pairs resulting in wins and losses will be presented by component of the composite endpoint determining the win or loss, by baseline bilirubin strata and overall. The number and percentage of pairs resulting in ties will be presented by baseline bilirubin strata and overall.

8.1.2.1.2. Calculation of the (1-α)% confidence interval

In the bootstrap method, the original data is sampled with replacement a large number of times, the win ratio is calculated for each sample, and an empirical distribution of the win ratio is determined from the total number of samples. Because the distribution of the win ratio is skewed, the (1-α)% CI for the population win ratio will be calculated by application of a bias-corrected method on the log-transformed win ratio, as recommended by Wang and Pocock (2016). The null hypothesis H_0 will be rejected in favor of the alternative H_a at the α% significant level if the (1-α)% CI does not contain 1.

The bootstrap method proceeds as follows:

Step 1: Within each baseline bilirubin strata, draw a random sample $S_{Placebo}$ of size $N_{Placebo}$ with replacement from the original placebo treatment group in the ITT popuation and a random sample S_{OCA} of N_{OCA} with replacement from original OCA treatment group in the ITT

popuation, where $N_{Placebo}$ and N_{OCA} are the total number of placebo and OCA patients in the ITT population, respectively.

Step 2: Perform the win ratio analysis on the sample ($S_{Placebo}$, S_{OCA}) and calculate the log win ratio (log Rw).

Step 3: Repeat steps 1 and 2 M times, where M is 5,000.

Step 4: Obtain (1-α)% CI for log R_w using the bias correction and acceleration (BCa) method and exponentiate these percentiles: these are the estimated limits of the (1-α)% CI for the win ratio.

8.1.2.1.3. Calculation of the Finkelstein-Schoenfeld Test Statistic

The stratified Finkelstein-Schoenfeld test is a generalization of the Gehan-Wilcoxon test which compares each pair of subjects i and j within strata k using a score:

$$u_{ij} \begin{cases} 1, \text{ if } i \text{ does "better than" } j \\ -1, \text{ if } i \text{ does "worse than" } j \\ 0, \text{ if it is indeterminate,} \end{cases}$$

where $i, j = 1, \ldots n_k$, with $i \neq j$ and n_k is the total number of subjects in the strata k. All subjects are compared pairwise with wins, loses, and ties determined as described in Section 8.1.2.1.1.

The "rank" for subject i is $U_i = \sum_{j \in A_k} u_{ij}$, where A_k is the set of indices of the n_k subjects in strata k. The test is based on $T = \sum_k \sum_{i \in A_k} U_i D_i$, where $D_i = 1$ for subjects in the treatment group and $D_i = 0$ for subjects in the placebo group, and variance

$V = \sum_k \frac{m_k(n_k - m_k)}{n_k(n_k - 1)} \left(\sum_{i \in A_k} U_i^2 \right)$, where m_k is the number of OCA-treated subjects in strata k.

The null hypothesis of this test is that none of the survival nor tiebreaker measures is affected by treatment, and the alternative is that at least one measure is improved in the treatment arm compared to the placebo arm, and is tested by comparing $T/V^{1/2}$ to the normal distribution.

8.2. Key Secondary Efficacy Endpoints

The key secondary efficacy endpoints are as follows:

- To assess the effect of OCA compared to placebo on time to first occurrence of death, liver transplant, MELD-Na score ≥15 if MELD-Na< 12 at baseline, MELD score ≥15 if MELD-Na ≥12 at baseline, uncontrolled or refractory ascites, portal hypertension syndromes (hepatorenal syndrome, portopulmonary syndrome, hepatopulmonary syndrome), or hospitalization for new onset or recurrence of variceal bleed, hepatic encephalopathy, spontaneous bacterial peritonitis, or bacterial empyema.

- Time to first occurrence of death, liver transplant, MELD ≥15, uncontrolled ascites, or hospitalization for new onset or recurrence of variceal bleed, hepatic encephalopathy, or spontaneous bacterial peritonitis. This will be used as the primary endpoint for European Medicines Agency (EMA) filing, and therefore all sensitivity analyses described in Section 8.1.2 will be performed for this endpoint, with the exception of the win ratio analysis. The endpoint described in Section 8.1 will be

considered the first key secondary endpoint for the EMA filing, followed by the remaining key secondary efficacy endpoints described in this section.

- Time to liver transplant or death (all-cause)

8.2.1.   Analyses of the Secondary Endpoints

The key secondary efficacy endpoints will compare OCA to placebo using the ITT Population. Data collected after the intercurrent events will be included (the treatment policy). Only adjudicated events will be included in the time-to-event analyses.

The time-to-event secondary efficacy analyses will compare OCA to placebo with respect to the secondary efficacy endpoints using the ITT Population. Only adjudicated events will be included in the time-to-event analyses. Time-to-event secondary efficacy analyses will use the same methodology as specified for the primary efficacy endpoint (Section 8.1.1).

8.2.2.   Sensitivity Analyses

The following sensitivity analyses will be carried out on the key secondary endpoints.

1. For key secondary time-to-event endpoints, the same analyses will be performed for PP population. Data collected after the intercurrent events will be included (treatment policy). No KM plot will be generated.

8.3.   Additional Secondary Efficacy Endpoints

8.3.1.   Time to Event Analyses

The following time-to-event secondary efficacy analyses will compare OCA versus placebo using the ITT population:

- Time to each relevant component of the primary efficacy endpoint

- Time to development of varix/varices

- Progression to cirrhosis

- Time to occurrence of HCC

- Time to liver-related death

- Time to liver-related death or liver transplant

- Time to liver-related death, liver transplant, or MELD score $\geq$15

All additional secondary time-to-event outcomes with the exception of death will be analyzed in the context of competing risks.

For the following secondary endpoints, liver transplant and death will be considered competing risks:

- Time to each relevant component of the primary efficacy endpoint, with the exception of death and liver-transplant

- Time to development of varix/varices

- Progression to cirrhosis

- Time to occurrence of HCC

For the following secondary endpoints, death will be considered a competing risk:

- Time to liver-transplant

For the following secondary endpoints, non-liver-related death will be considered a competing risk:

- Time to liver-related death

- Time to liver-related death or liver transplant

- Time to liver-related death, liver transplant, or MELD score ≥15

Competing risk outcomes will be compared and summarized through a cumulative incidence function (CIF). Gray's (1988) non-parametric test will be used to compare the CIFs of the two treatment groups. Hazard ratios and their 95% confidence intervals will be computed following the Fine and Gray model (1999). Fine and Gray's hazard model can be fit in SAS version 9.4 and higher by specifying eventcode option in PROC PHREG.

Progression to cirrhosis will be assessed in the subset of subjects considered noncirrhotic at Baseline using available medical history, clinical, and laboratory assessments as well as baseline transient elastography (TE), where available. Specifically, subjects with no clinical, laboratory, or radiological evidence of cirrhosis at baseline and/or a TE liver stiffness of <16.9 kPa (Corpechot 2012) will be considered noncirrhotic. Progression to cirrhosis will be assessed by the development of clinical, laboratory, or radiological evidence of cirrhosis and/or when liver stiffness increases to ≥16.9 kPa during the study in which case histological progression to cirrhosis should be confirmed by biopsy unless not medically indicated.

The percentage of noncirrhotic subjects progressing to cirrhosis during the treatment period will also be summarized by treatment group and compared using a Cochran-Mantel-Haenszel (CMH) test stratified by the randomization stratification factor.

8.3.2.   Liver Biochemistry Analyses

Analyses of changes in liver biochemistry (gamma-glutamyl transferase (GGT), ALT, AST, conjugated bilirubin, albumin, and INR) at each visit will be compared between treatment groups using a restricted maximum likelihood based mixed-effect model repeated measures (MMRM) with treatment, baseline, visit, visit by treatment interaction and randomization stratification factors to be included in the model. Visits through Month 24 will be included. An unstructured covariance model will be used. If the computational algorithm fails to converge, the following structures will be tested in the order: Toeplitz, AR(1) and compound symmetry (CS). If the model still does not converge, the stratification factors will be dropped from the model one at a time (UDCA followed by bilirubin).

Change from Baseline over time will be analyzed using analysis of covariance (ANCOVA) with change from Baseline as the dependent variable including treatment group and randomization stratification factor as fixed effects and the Baseline values as a covariate. Descriptive statistics of the laboratory values will be summarized by treatment group. The results, change from

Baseline, and percentage change from Baseline values, as well as estimates of least-square means, standard errors, and 95% CIs, will be presented by treatment group. Estimates of the mean difference between treatment groups, the standard error of the difference, and 95% CI of the difference will be presented.

8.3.3. Prognostic Endpoints and Markers of Inflammation

Change and percentage change from Baseline for MELD score, MELD-Na score, Child-Pugh score, Mayo Risk Score (MRS), IgM, C-reactive protein (CRP), tumor necrosis factor-α (TNF-α), FGF-19, cytokeratin-18 (CK-18), C4, and enhanced liver fibrosis (ELF) score will be summarized and analyzed using the Wilcoxon Rank Sum Test to compare groups at annual timepoints. The median differences and 95% CI of the median differences between treatment groups will be constructed using Hodges-Lehmann estimate. If the median differences between treatments are not approximately symmetrical around the median, the Sign Test may be used.

Values below or above the quantifiable limits will be handled as follows. Quantitative laboratory tests containing less than (<) and greater than (>) symbols are test results that are below and above quantifiable limits, respectively. In order to retain these values for analysis purpose, the following imputations will be done within the analysis datasets:

For laboratory test results that are below the quantifiable limit:

Imputed laboratory results = (numeric portion of the result) × 0.9.

For laboratory test results that are above the quantifiable limit:

Imputed laboratory results = (numeric portion of the result) × 1.1

8.4. Exploratory Efficacy Analyses

8.4.1. Acute-on-Chronic Liver Failure (ACLF)

Shift tables (i.e., status at baseline versus status at follow-up visit) will be presented for the Chronic Liver Failure Consortium Acute Decompensation score (CLIF-C ADs) using the categories: CLIF-C ADs ≤ 45, 45 < CLIF-C ADs ≤ 60, and CLIF-C ADs > 60. The number and percentage of subjects in each category will be presented by treatment group and visit.

8.4.2. Biochemical Responder Analyses

The percentage of subjects that meet the following criteria of a responder based on each of the definitions below will be summarized by treatment group and at each study visit:

- Decrease in ALP of ≥15% and ≥40% from Baseline

- ALP ≤ ULN

- ALP ≤3x ULN and AST ≤2x ULN and total bilirubin ≤ULN (Corpechot 2008)

- ALP ≤1.5x ULN and AST ≤1.5x ULN and total bilirubin ≤ULN (Corpechot 2011)

- ALP ≤1.67x ULN and total bilirubin ≤ULN (Momah 2012)

- Normal bilirubin (values ≤ULN) and normal albumin (values ≥lower limit of normal) (Kuiper 2009)

- ALP ≤1.76x ULN (Kumagi 2010)

Responder analyses will compare treatment groups using a CMH test stratified by the randomization stratification factor. Missing values will be considered as a non-responder.

8.4.3.   Time-to-Event by Biochemical Response

Additional efficacy analyses of correlation of time-to-event outcomes and biochemical response will be run as supportive information for primary endpoint evaluation and in support of FDA PMR 3057-3 as described below.

A biochemical responder is defined as a subject who attains an ALP <1.67 times the upper limit of normal (ULN) and total bilirubin ≤ULN and an ALP decrease of ≥15% from Baseline at their 12-month visit. Subjects missing values are considered non-responders. The number and percent of subjects censored and with events and descriptive statistics for the time-to-event will be presented by biochemical response. KM estimates of the distribution of the time-to-event using the primary composite endpoint will be tabulated and graphed by treatment group and biochemical responder status. KM tabulations will present time to event in days; KM graphs will present time to event in months.

Biochemical responders and non-responders will be compared using a log rank test stratified by treatment group. The proportionality of hazards will be assessed using Schoenfeld residuals and the hazard ratio and 95% CI will be estimated based on a Cox regression model stratified by treatment group.

The analysis will be repeated using the following alternative definitions of biochemical responder:

- ALP ≤2.0 x ULN at their 12-month visit

- Total bilirubin <1 x ULN at their 12-month visit

For placebo subjects, OCA-treated subjects, and all ITT subjects, additional KM estimates of the distribution of the time-to-event will be graphed by biochemical response for the combined ALP and bilirubin categories:

1. ALP ≤2.0 x ULN and Total bilirubin <1 x ULN at their 12-month visit

2. ALP >2.0 x ULN and Total bilirubin <1 x ULN at their 12-month visit

3. ALP ≤2.0 x ULN and Total bilirubin ≥1 x ULN at their 12-month visit

4. ALP >2.0 x ULN and Total bilirubin ≥1 x ULN at their 12-month visit

8.4.4.   Liver Histology

A paired biopsy sub-study was recommended by Regulatory Authorities, to further assess clinical outcomes in terms of histological progression to cirrhosis and is described in Protocol Addendum 2. While biopsies were collected, no paired data exists; no analysis will be run.

8.5.   Interim Analyses

A single interim analysis was conducted when 63% of the total events were reached. The interim analysis was conducted and reviewed by the Data Monitoring Committee (DMC) according to the Interim Analysis Statistical Analysis Plan. No further interim analysis for superiority or futility is planned for this study. This is a deviation from the protocol. Further details are described in Section 14.1.

8.6.   Multiple Comparisons/Multiplicity

An interim analysis was conducted after 80 events were observed. A group sequential design using the O'Brien-Fleming type alpha-spending function (Lan 1983, Reboussin 2000) was used to control the overall two-sided 0.05 significance level that is at an information fraction of 0.63 of the total of 127 events, corresponding to an alpha spend of 0.009. The primary analysis will be conducted at 2-sided 0.041 significance level. No interim analysis will be performed based on the updated endpoints and methodology. The primary endpoint analysis for European Medicines Agency (EMA) filing will be conducted at 2-sided 0.041 significance level.

For the final analysis, the hypothesis testing of key secondary analyses will compare placebo and OCA (the key secondary endpoints are stated in Section 8.2), provided that the primary efficacy endpoint comparison is statistically significant in favor of OCA. If the primary comparison is not statistically significant then comparison of the key secondary endpoints will be considered as supportive or supplemental. If the primary comparison is statistically significant, the key secondary endpoints will be tested hierarchically in the order described in Section 8.2 at a 2-sided 0.041 significance level. All other secondary endpoints will be tested at nominal 0.041 alpha level.

8.7.   Subgroup Analyses of Efficacy

The primary and all key secondary efficacy endpoints (Section 8.2), liver biochemistry (Section 8.3.2), and prognostic endpoints and markers of inflammation endpoints (Section 8.3.3) will be analyzed for subject subgroups based on the ITT population. Sensitivity analyses will not be run for subgroup analyses. Subgroups will be assessed at Baseline. Baseline subgroups of interest are as follows:

- Baseline disease stage:
  - On-label per United States Prescribing Information (USPI). The on-label subgroup includes ITT subjects who had not experienced clinically evidence portal hypertension (CSPH) or decompensation at baseline.
    - A subject is defined to have CSPH at baseline if they had any of the following: a previous history of portal hypertension includes procedures for TIPS, sclerotherapy, ligation, HVPG measurement >10 mmHG, paracentesis, thoracentesis (due to hepatic hydrothorax); previous history or baseline observation of preferred terms for presence of collateral circulations secondary to CSPH, GI bleeding due to varices or portal hypertension, gastroesophageal varices without bleeding and portal hypertensive gastropathy, ascites, hepatopulmonary syndrome,

> hepatorenal syndrome, portopulmonary hypertension, hepatic encephalopathy; platelets $<150 \times 10^9$/L with splenomegaly and/or with transient elastography >15 kPa.

- A subject is defined to have decompensation at baseline if they had any of the following: a Child-Pugh score of B or C; previous history or baseline observation of preferred terms for gastric variceal or esophageal variceal bleeding, ascites, hepatic hydrothorax, spontaneous bacterial peritonitis, hepatic encephalopathy, hepatorenal / hepatopulmonary / portopulmonary syndrome, prior TIPPS or other peritoneovenous shunt.

- Compensated, non-cirrhosis or cirrhosis. The pre-decompensation subgroup includes ITT subjects who had not experienced decompensation at baseline.

- Decompensated. The decompensated subgroup includes ITT subjects who had experienced decompensation at baseline.

- Contraindicated per USPI. The contraindicated subgroup includes ITT subjects who had experienced CSPH and/or decompensation at baseline.

The primary efficacy endpoint and key secondary endpoints will also be assessed for the following baseline subgroups of interest if there are a sufficient number of subjects in each group (eg, >5 subjects per group):

- Sex: male, female

- Age categories (at time of informed consent): <65 years, ≥65 years

- Age categories at PBC Diagnosis (at time of informed consent): <50 years, ≥50 years

- Race: white, non-white

- Ethnicity: Hispanic, non-Hispanic

- Baseline bilirubin level: >ULN, ≤ULN

- Baseline use of UDCA: Yes, No

The primary efficacy endpoint will also be assessed across the spectrum of PBC disease stage as defined by Rotterdam criteria (Kuiper 2009)

- Early (normal albumin and normal bilirubin)

- Moderate (abnormal albumin or abnormal bilirubin)

- Advanced (abnormal albumin and abnormal bilirubin)

If a strong correlation between biochemistry and clinical outcomes using the ITT population is observed, the biochemical improvement in the subpopulations (baseline disease stage, Rotterdam disease severity, and monotherapy) will be further assessed.

9.   SAFETY ANALYSES

Safety analyses include exposure to study treatment, treatment-emergent adverse events (TEAEs), clinical laboratory parameters, electrocardiograms (ECGs), vital signs, and any abnormal findings observed during physical examinations after study enrollment and through discontinuation of study treatment or within 30 days after the discontinuation of study treatment. Data after a patient begins commercial OCALIVA will be excluded. Safety analyses will be performed using the Safety Population including by treatment groups. No inferential comparison of safety endpoints will be performed, unless otherwise specified.

9.1.   Exposure to Study Treatment

The extent of exposure will be summarized using descriptive statistics. Duration of exposure (days) to IP will be calculated as follows:

> Exposure to investigational product = Date of last IP dose – Date of first IP dose + 1 – Total duration of temporary IP discontinuation.

The duration of each incidence of temporary investigational product discontinuation will be calculated as follows:

> Duration of temporary discontinuation of investigational product = Date of restart of IP – Date of temporary discontinuation of IP + 1.

The total duration of temporary IP discontinuation is the sum duration of temporary discontinuation of IP over each incidence of discontinuation.

Total subject exposure to IP will be calculated by adding the doses taken by a subject during the study and will be summarized using descriptive statistics.

A summary of subjects who had an increase in dose and who had a decrease in dose at least once during the study will be provided in terms of frequency count (n) and percentages (%).

Dose and frequency will be modified for subjects with cirrhosis (including subjects progressing to cirrhosis during the study) and classified as CP-B or CP-C. Therefore, any adjustments, interruptions, discontinuation, and any rechallenge of IP will be summarized.

Subject's overall compliance (%) with IP will be calculated as follows:

> 100 × (number of tablets consumed during study / number of tablets expected to be consumed during study)

where

- number of tablets expected to be consumed during study = number of dispensed tablets during the study

and

- number of tablets consumed during study = number of tablets dispensed – number of tablets returned.

Subject compliance with IP will be summarized by treatment group using descriptive statistics. Percent compliance will be summarized separately for subjects who completed the study and

those who withdrew early so as to distinguish between those subjects who were compliant throughout the entirety of the study versus those who were compliant until they withdrew from the study.

9.2.   Adverse Events

An adverse event (AE) is defined as any untoward medical occurrence associated with the use of the investigational product in humans, whether or not considered related to investigational product. An AE (also referred to as an adverse experience) can be any unfavorable and unintended sign (eg, an abnormal laboratory finding), symptom, or disease temporally associated with any use of the investigational product, without any judgment about causality and irrespective of route of administration, formulation, or dose, including an overdose.

AEs include, but are not limited to: (1) a worsening or change in nature, severity, or frequency of condition(s) present at the start of the study; (2) subject deterioration due to primary illness; (3) intercurrent illness; and (4) drug interaction. For reporting purposes, pregnancy is not considered an AE.

For subjects who enrolled into the study, AEs will be mapped to PTs and SOCs using MedDRA dictionary version 23.1 (or later version). Subjects experiencing the same event more than once will be counted only once at the most severe grade and the closest relationship to study treatment.

9.2.1.   Treatment-Emergent Adverse Events

A TEAE is any event not present prior to the initiation of the IP or any event already present that worsens in either intensity or frequency following exposure to the IP or within 30 days after the discontinuation of the study treatment.

9.2.2.   Serious Adverse Events (SAE)

An AE is considered "serious" if, in the view of either the Investigator or Sponsor, it results in any of the following outcomes:

- Death

- Is immediately life threatening

- Requires in-subject hospitalization or prolongation of existing hospitalization

- Results in persistent or significant disability or incapacity

- Results in a congenital abnormality or birth defect

- Is an important medical event that may jeopardize the subject or may require medical intervention to prevent one of the outcomes listed above

Events not considered to be SAEs are hospitalizations for:

- Routine monitoring of the studied indication and not associated with any deterioration in condition or AE

- Elective treatment for a pre-existing condition that did not worsen

- Respite care or observation when there is no AE associated with the hospitalization

9.2.3.   Adverse Events of Special Interest (AESI)

The following decompensation events are adverse events of special interest:

- Hepatic decompensation/disorder events
  - Hepatic Disorders SMQ, excluding the following sub-SMQs:
    - Alcohol related
    - Congenital, familial, neonatal, and genetic disorders of the liver
    - Liver infections
    - Pregnancy-related hepatic disorders
- Cholecystitis/Cholelithiasis
  - Defined by Gallbladder related disorders narrow SMQ and Gallstone related disorders narrow SMQ
- Renal
  - Defined as events included in the following broad SMQs: acute renal failure, chronic kidney disease, proteinuria, renovascular disorders, or tubulointerstitial diseases
- Cardiovascular
  - Defined by the embolic and thrombotic events broad SMQ, ischaemic heart disease broad SMQ, or central nervous system vascular disorders narrow SMQ.
  - These events may be reassessed stratified by statin (or other lipid lowering medication) usage
- Dyslipidemia
  - Defined by the Dyslipidemia SMQ
  - These events may be reassessed stratified by statin (or other lipid lowering medication) usage
- Pruritus
  - Defined as any preferred term within the Pruritus Not Elsewhere Classified (NEC) high level term or any preferred term including "prur"

9.2.4.   TEAE Analyses

The number of events and counts and percentages of unique subjects experiencing a TEAE and TEAEs by PTs and SOC in the following categories will be presented for the Safety Population:

- TEAEs

- TEAEs by highest severity (mild, moderate, severe): At each level of subject summarization, a subject is classified according to the highest severity if the subject reported 1 or more events. AEs with missing severity will be considered severe for this summary.

- Serious AEs

- Adverse events of special interest

- AESIs by highest severity (mild, moderate, severe): At each level of subject summarization, a subject is classified according to the highest severity if the subject reported 1 or more events. AEs with missing severity will be considered severe for this summary.

- TEAEs leading to study treatment discontinuation: This is a subset of the AEs where action taken with study treatment is checked as "Drug withdrawn" or where "Subject Discontinued from Study" is checked.

- TEAEs leading to study treatment interruption: This is a subset of the AEs where action taken with study treatment is checked as "Drug interrupted".

- TEAEs leading to study treatment dose reduction: This is a subset of the AEs where action taken with study treatment is checked as "Dose reduced".

- Deaths: This is a subset of the AEs where "Fatal" is checked.

- Subject incidence of TEAEs by MedDRA SOC, PT, and closest relationship to investigational product (Related/Not Related). Related AEs are those with relationships reported by the Investigator as "Definite," "Probable," or "Possible," and unrelated AEs are those with relationships reported as "Unlikely" or "Not Related." At each level of subject summarization, a subject is classified according to the closest relationship if the subject reported 1 or more events. AEs with a missing relationship will be considered related for this summary.

Summaries will be sorted in order of decreasing frequency in the treatment arm. Summaries will also be presented by MedDRA SOC and PT, sorted in alphabetic order of SOC and then in order of decreasing frequency in the treatment arm.

Exposure-adjusted incidence of TEAEs will also be provided by SOC and PT. Each individual subject year on study will be derived as the last known date during the study minus the first dose date plus 1 divided by 365.25 days/year.

Exposure-adjusted incidence rates will additionally be provided by SOC and PT for "on-risk" AEs. An on-risk adverse event (ORAE) is any event not present prior to the initiation of the IP or any event already present that worsens in either intensity or frequency following exposure to the IP or within 30 days after the discontinuation of the regularly-scheduled study visits. As some subjects may be on commercial OCALIVA after discontinuing study treatment and prior to discontinuing regularly scheduled study visits, ORAE events will be summarized prior to initiation of commercial OCALIVA.

9.3.   Adjudicated Cardiovascular Events

Major Adverse Cardiovascular Events (MACE) (defined as death, myocardial infarction, and stroke) are included in the Cardiovascular Adjudication Committee Charter for adjudication. The adjudication of these events will be handled separately from the adjudication of events for assessment of the primary clinical composite outcome endpoint in this study.

Adjudicated MACE is a time-to-event endpoint and will be summarized for the Safety Population.

OCA and placebo treatment groups will be compared using the log-rank test stratified by the randomization stratification factors. In addition, the hazard ratio and 95% CI will be determined based on a Cox regression model stratified by randomization strata to estimate the magnitude of the effect.

The number and percent of subjects censored and with events will be presented. Descriptive statistics will be presented for the time-to-event.

Kaplan-Meier (KM) estimates of the distribution of the time-to-event will be tabulated and graphed by treatment group. The tabulation will include the KM estimate of the 25th, 50th (median), and 75th percentiles and corresponding 95% CIs, where the percentiles can be estimated. KM tabulations will present time to event in days; KM graphs will present time to event in months. The hazard ratio and 95% CI will be estimated based on a Cox regression model stratified by randomization strata.

Only adjudicated events will be included in analyses. Subjects without any documentation of events will be censored at the date of last contact. For subjects with more than one event, the earliest of the event dates will be used.

9.4.   Clinical Laboratory Evaluations

Laboratory parameters will be summarized in both conventional units and the standard international (SI) system of units using the Safety Population by treatment group and descriptive statistics at Baseline and at each scheduled study visit. Only central laboratory data will be used.

9.4.1.   Hematology and Coagulation

Descriptive statistics will be used to summarize the results and change and percentage change from Baseline to each on-study evaluation visit for hemoglobin, hematocrit, white blood cell with differential (lymphocytes, monocytes, eosinophils, basophils, neutrophils), platelets, red blood cell count [including mean corpuscular volume, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration], prothrombin time (PT) and INR, and partial thromboplastin time (PTT).

9.4.2.   Chemistry

Descriptive statistics will be used to summarize the results and change and percentage change from Baseline to each on-study evaluation visit for albumin, blood urea nitrogen, creatinine, total bilirubin, unconjugated (indirect) bilirubin, conjugated (direct) bilirubin, GGT, AST, ALT, ALP, electrolytes (calcium, chloride, potassium, sodium), magnesium, phosphorous, bicarbonate,

glucose, total protein, total cholesterol, low-density lipoprotein, high-density lipoprotein, very low density lipoprotein (VLDL), triglycerides, free fatty acids.

9.5.   Vital Signs

Descriptive statistics will be used to summarize the results and change and percentage change from Baseline to each on-study evaluation visit for systolic blood pressure, diastolic blood pressure, heart rate, and temperature.

9.6.   Electrocardiograms

The central read ECG data will be analyzed based on methodology recommended in the ICH E14 guideline, The Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Nonantiarrhythmic Drugs.

Overall interpretation results for ECGs and the investigator interpretation results are collected as normal, abnormal not clinically significant (NCS), and abnormal clinically significant (CS). Subjects whose interpretations shift from normal to abnormal (CS or NCS) will be listed separately, including description of the abnormality and any associated comments.

9.7.   Concomitant Medications

A concomitant medication is any medication that was started after the initiation of IP.

Concomitant medications will be mapped to ATC class and PT using WHO Drug Dictionary version Global B3, March 2020 (or later version) and summarized by ATC class, PT, and treatment in the Safety Population. Summaries will be ordered by descending order of incidence of ATC class and preferred term within each ATC class.

9.8.   Subgroup Analyses of Safety

Demographics and baseline disease characteristics (Section 6), exposure to study treatment (Section 9.1) and the adverse events (Section 9.2) analyses listed below will be presented for the baseline disease stage subgroup described in Section 8.7.

The number of events and counts and percentages of unique subjects experiencing a TEAE and TEAEs by PTs and SOC in the following categories will be presented:

- TEAEs
- Serious AEs
- Adverse events of special interest
- TEAEs leading to study treatment discontinuation
- Deaths

Exposure-adjusted incidence of TEAEs will also be provided by SOC and PT.

10.   PHARMACOKINETIC ANALYSES

Analyses of pharmacokinetic outcomes will be specified in a separate report.

11.   PHARMACOECONOMIC AND HEALTH OUTCOMES ANALYSES

Analyses of pharmacoeconomic and health outcomes will be specified in a separate report.

12.   PHARMACODYNAMIC ANALYSES

Analyses of pharmacodynamic outcomes will be specified in a separate report.

13.   EXTERNAL CONTROL COMPARISON

The exploratory analyses comparing OCA to external control data will be performed as post-hoc analyses and will be described in a separate SAP.

14.   DEVIATIONS FROM PROTOCOL

14.1.   Primary Efficacy Endpoint and Analysis

Due in part to high discontinuation rates and requests from sites to close, the decision was made to stop 747-302 early, at approximately 67% information (84 adjudicated Protocol Version 6 endpoints observed). To increase the number of endpoint events and power to detect a difference, the primary efficacy endpoint has been expanded to include:

- Portal hypertension syndromes (hepatorenal, portopulmonary, or hepatopulmonary)

- Progression to decompensated liver disease (for subjects without decompensation at baseline)

- Progression to clinical evidence of portal hypertension without decompensation (for subjects without decompensation or clinical evidence of portal hypertension at baseline).

The primary efficacy analysis remains the stratified log rank test. With the inclusion of additional endpoints that occur earlier in the course of the disease, the log rank test is limited to each subject's first event, which is likely to be an outcome of lesser clinical importance. A win ratio analysis was added as a sensitivity analysis, to allow for a hierarchical comparison of the composite endpoints. The win ratio analysis allows for the comparison of more severe events such as death and liver transplant to be prioritized over the comparison of earlier-stage events such as progression to clinical evidence of portal hypertension.

14.2.   Number of Interim Analyses

Although the protocol did originally plan for two interim analyses, one at ~50% (64 events) and the other at ~75% (96 events) information fraction, the first interim analysis was postponed to occur after regulatory authority review of the interim SAP. As a result, the first interim analysis occurred after the accrual of approximately 65% (83 events) of the required total clinical outcome events. The Sponsor proposed and the regulatory authority agreed that the second interim analysis will be omitted. O'Brien-Fleming Boundaries were used to determine the criteria for superiority in the interim analysis. Table 2 provide the relevant estimates both for the original two-interim analyses and revised single-interim analysis scenarios.

Table 2:   Stopping Criteria Based on O'Brien-Fleming Boundaries

Type 1 Error	Power	Information Fraction	Number of Clinical Events	Incremental α Spend	Cumulative α Spend	Reject for Efficacy if Hazard Ratio
Two Interim Analyses (Originally Planned)						
		50%	64	0.003	0.003	≤0.473
0.05	0.80	75%	96	0.018	0.019	≤0.611
		End of Study (100%)	127	0.044	0.050	≤0.695
Single Interim Analysis						
		65%	83	0.011	0.011	≤0.564
0.05	0.80	End of Study (100%)	127	0.047	0.050	≤0.697

15.   REFERENCES

Angeli P, Garcia-Tsao G, Nadim MK, Parikh CR. News in pathophysiology, definition and classification of hepatorenal syndrome: a step beyond the International Club of Ascites (ICA) consensus document. Journal of hepatology. 2019 Oct 1;71(4):811-22.

Corpechot C, Abenavoli L, Rabahi N, et al. Biochemical response to ursodeoxycholic acid and long term prognosis in primary biliary cirrhosis. Hepatology. 2008;48(3):871-877.

Corpechot C, Carrat F, Poujol-Robert A, et al. Noninvasive elastography-based assessment of liver fibrosis progression and prognosis in primary biliary cirrhosis. Hepatology. 2012;56(1):198-208.

Corpechot C, Chazouillères O, Poupon R. Early primary biliary cirrhosis: Biochemical response to treatment and prediction of long-term outcome. J Hepatol. 2011;55:1361–1367.

Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. Journal of the American statistical association. 1999 Jun 1;94(446):496-509.

Finkelstein D.M. and Schoenfeld D.A., Combining mortality and longitudinal measures in clinical trials. Statistics in Medicine, 1999; 18(11): 1341-1354.

Gray RJ. A class of K-sample tests for comparing the cumulative incidence of a competing risk. The Annals of statistics. 1988 Sep 1:1141-54.

Guidance for Industry E14: *The Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-antiarrhythmic Drugs*, U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), October 2005.

Kuiper EM, Hansen BE, de Vries RA, et al. Improved prognosis of patients with primary biliary cirrhosis that have a biochemical response to ursodeoxycholic acid. Gastroenterology. 2009;136(4):1281-7.

Kumagi T, Guindi M, Fischer SE, et al. Baseline ductopenia and treatment response predict long-term histological progression in primary biliary cirrhosis. Am J Gastroenterol. 2010;105(10):2186-94.

Lan KKG and DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika. 1983;70:659-663.

Momah N, Silveira MG, Jorgensen R, et al. Optimizing biochemical markers as endpoints for clinical trials in primary biliary cirrhosis. Liver Int. 2012;32(5):790-795.

Pocock SJ, Ariti CA, Collier TJ,Wang D. The Win Ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *Eur Heart J*. 2012;33(2):176-82.

Reboussin DM, DeMets, DL, Kim KM, et al. Computations for Group Sequential Boundaries Using the Lan-DeMets Spending Function Method. Controlled Clin Trials. 2000;21(3):190-207

Robins JM. Association, Causation, and Marginal Structural Models. Synthese. 1999 Jan 1:151-179.

Wang D, Pocock S. A win ratio approach to comparing continuous non-normal outcomes in clinical trials. Pharmaceutical statistics. 2016 May;15(3):238-45.

APPENDIX A. GENERAL ANALYSIS RULES AND DATA CONVENTIONS

The general analysis rules and data conventions described herein that are pertinent to the interim analysis will be followed.

Individual subject data obtained from electronic case report forms (eCRFs), central laboratories, external sources, and any derived data will be presented in data listings by subject. All data listings that contain an evaluation date will contain a relative study day. Pre-treatment and on-treatment study days are numbered relative to the day of the first dose of investigational product which is designated as Day 1. The preceding day is Day -1, the day before that is Day -2, etc. The last day of investigational product is designated with an "L" (eg, Day 14L). Post-treatment study days are numbered relative to the last dose and are designated as Day 1P, Day 2P, etc.

All output will be incorporated into Microsoft Word rich text format (.rtf) files, sorted and labeled according to the ICH recommendations, and formatted to the appropriate page size(s).

Tabulations will be produced for appropriate demographic, Baseline, efficacy, and safety parameters.

For categorical variables, summary tabulations of the number and percentage of subjects within each category (with a category for missing data) of the parameter will be presented. Percentage calculations will be based on non-missing data, unless otherwise specified. Percentages are rounded to 1 decimal place, unless otherwise specified.

For continuous variables, the number of subjects, mean, standard deviation (SD), standard error of the mean (SEM), median, first and third quartiles (Q1 and Q3), minimum, and maximum values will be presented. Other summaries (eg, quartiles, 5%, 95% intervals) may be used as appropriate. The precision of summary statistics, unless otherwise specified, will be as follows: mean and median to 1 more decimal place than the raw data, and SD and SEM to 2 decimal places more than the raw data. In general, the decimal places should not exceed 3 decimal places unless appropriate. Confidence intervals (CIs) will be provided and will be rounded to 1 decimal place, unless otherwise specified, in the table and listing shell.

Time-to-event data will be summarized using Kaplan-Meier methodology using 25th, 50th (median), and 75th percentiles with associated 2-sided 95% CIs, as well as percentage of censored observations.

All statistical tests comparing groups will be conducted at the 2-sided, 0.05 level of significance, unless otherwise specified (eg, the primary efficacy analysis, Section 8.6). Summary statistics for each treatment group will be presented, as well as 95% CIs comparing groups.

Baseline Definitions

The Baseline value for statistical analyses of quantitative parameters is defined as the mean of all available study evaluations after the subject signs the informed consent and prior to or on the first administration of investigational product, unless otherwise specified. If there is only one evaluation prior to the first administration of investigational product then the available data from this evaluation will be used as the Baseline value.

The baseline value for analyses of lipid parameters is defined as the last fasted evaluation prior to the first administration of IP.

The baseline value for analyses of qualitative parameters (eg, normal/abnormal) is defined as the last evaluation prior to the first administration of IP.

Baseline values defined above will not change regardless if a subject stops taking investigational product and begins taking commercially marketed OCA.

Visit Windows

Visit windows, as provided below, will be established relative to an individual's Day 0, the date of first administration of investigational product. Analyses will use visit windows defined by relative day, overriding CRF designated visit numbers.

OCA Therapy 3-Month Interval Visit Windows (Days)

		Analysis Window Study Day [a]	
Months	**Target Study Day [a]**	**Low**	**High**
Day 0	1	Closest visit to Day 1, prior to first IP dose [b]	
Month 1	30	2	46
Month 3	90	47	137
Month 6	183	137	227
Month 9	274	228	318
Month 12	365	319	410
…	…	…	…
Month n [b]	round(365.25/12*n)	round(365.25/12*n) – 46	round(365.25/12*(n+3)) – 47

Note: Follow-up visits are not included in visits windows.
[a] Study day will be calculated from first dose date.
[b] For n in three month increments (15, 18, 21, …).

Partial Dates

If only a partial date is available and is required for calculation, the following standards will be applied:

- Diagnostic Date (eg, PBC diagnostic date)
  - For missing day only – Day will be imputed as the first day of the month (ie, 1).
  - For missing day and month – Day and month will be imputed as the first day of the year (ie, 1 January).
- Start Dates (eg, event date, AE onset date, or start date of medication)

   – For missing start day only – Day will be imputed as the first day of the month (ie, 1) with the following exception: if the partial date falls in the same month and year as the date being used in the calculation (eg, first dose date, informed consent date), then the partial date will be imputed to equal the date being used for the calculation.

   – For missing start day and month – Day and month will be imputed as the first day of the year (ie, 1 January) with the following exception: if the partial date falls in the same year as the date being used in the calculation (eg, first dose date, informed consent date), then the partial date will be imputed to equal the date being used for the calculation.

   – Imputed start dates must be prior to the stop date.

- Stop Dates (eg, AE resolution date or stop date of medication)

   – For missing stop day only – Day will be imputed as the last day of the month (ie, 28, 29, 30, or 31).

   – For missing stop day and month – Day and month will be imputed as the last day of the year (ie, 31 December).

   – Imputed stop dates must be on or after the start date.

All data recorded on the case report form will be included in data listings that will accompany the clinical study report.

Data Conventions

The precision of original measurements will be maintained in summaries, when possible.

Means, medians, SEMs, and SDs will be presented with an increased level of precision, where means and medians will be presented to one more decimal place than the raw data, and the SEMs and SDs will be presented to two more decimal places than the raw data. In general, the decimal places should not exceed three decimal places, unless appropriate.

For tables where rounding is required, rounding will be done to the nearest round-off unit. For example, when rounding to the nearest integer, values $\geq$XX.5 will be rounded up to XX+1 (eg, 97.5 will round up to 98), while values <XX.5 will be rounded down to XX (eg, 97.4 will round down to 97).

Percentages based on frequency counts will be based on available data, and denominators will generally exclude missing values, unless otherwise stated. For frequency counts of categorical variables, categories whose counts are zero will be displayed for the sake of completeness. For example, if none of the subjects discontinue due to "lost to follow-up," this reason will be included in the table with a count of 0. Percentages based on frequency counts will be presented as a whole number (no decimal places), and values less than 1% will be presented as "<1%." Values less than 100% but that round up from 99.5% to 100% will be presented as ">99%."

Quantitative laboratory tests containing less than (<) and greater than (>) symbols are test results that are below and above quantifiable limits, respectively. In order to retain these values for analysis purpose, the following imputations will be done within the analysis datasets:

For laboratory test results that are below the quantifiable limit:

Imputed laboratory results = (numeric portion of the result) x 0.9.

For laboratory test results that are above the quantifiable limit:

Imputed laboratory results = (numeric portion of the result) x 1.1.

For the purpose of tabulations, the unscheduled post-baseline values generally will be excluded from summary tables, but will be included in data listings. Unscheduled visits will be considered for analyses of shift from baseline to worst value (low-normal-high).

Standard Calculations

Variables requiring calculation will be derived using the following formulas:

- First dose date is defined as the day of first dose of IP received after randomization

- Last dose date is defined as day of the last dose of IP

- Time to event – The time to an event will be calculated in days as the date of the first occurrence of the event minus the date of first investigational product administration plus 1

- Days – A duration expressed in days between one date (*date1*) and another later date (*date2*) will be calculated using the following formulas:

    duration in days = date2 – date1 + 1, where date1 ≥first dose date

    duration in days = date2 – date1, where date1 <first dose date

- Months – A duration expressed in months is calculated as the number of days divided by 365.25/12 (approximately 30.4)

- Years – A duration expressed in years between one date (*date1*) and another date (*date2*) is calculated using the following formulas:

    duration in years = (date2 – date1 + 1)/365.25, where date1 ≥first dose date

    duration in years = (date2 – date1)/365.25, where date1<first dose date

- Age – Age is calculated as the number of years from the date of birth (*DOB*) to the specified date, eg, date of informed consent (*DOIC*). If the month of DOIC <month of DOB or the month of DOIC = DOB and the day of DOIC <day of DOB, then the following formula is used:

  age (years) = year of DOIC – year of DOB – 1.

  Otherwise, the following formula is used:

  age (years) = year of DOIC – year of DOB.

  If only year is provided in DOB, then July 1 will be used for the month and day.

- Change from Baseline – Change from Baseline will be calculated as:

  Change = post Baseline value – Baseline value

- Percentage change from Baseline – Change from Baseline will be calculated as:

  Percentage change from Baseline = ([post Baseline value – Baseline value] / Baseline value) × 100

- MELD score is derived using the following formula:

  MELD = 3.78 × ln[serum bilirubin (mg/dL)] + 11.2 × ln[INR] + 9.57 × ln[serum creatinine (mg/dL)] + 6.43 × aetiology(0: cholestatic or alcoholic, 1: otherwise).

  Laboratory values less than 1.0 are set to 1.0 when computing the MELD score.

- MELD-Na score is derived using the following formula:

  MELD-Na = MELD + 1.32 × (137-[adjusted sodium (mmol/L)]) – [0.033 × MELD × (137-[adjusted sodium (mmol/L)])],

  where adjusted sodium is equal to

  - [serum sodium (mmol/L)] + 0.024 × ([serum glucose (mg/dL)] -100)

    when serum glucose >100 mg/dL

  - [serum sodium (mmol/L)]

    when serum glucose ≤100 mg/dL

  and adjusted sodium values less than 125 mmol/L are set to 125, and values greater than 137 mmol/L are set to 137.

- Child Pugh classification (Noncirrhotic/A/B/C) is defined as follows: Subjects deemed noncirrhotic at baseline by the PI are captured as "noncirrhotic" regardless of baseline Child-Pugh score and those without a cirrhosis assessment or who were deemed cirrhotic at baseline are summarized by their baseline (Screening or Day 0) Child-Pugh category.

- Rotterdam Criteria (Mild/Moderate/Severe) is defined as follows:

  Mild - Total Bilirubin ≤ ULN and Albumin ≥ LLN

  Moderate - Total Bilirubin > ULN and Albumin ≥ LLN or

Total Bilirubin ≤ ULN and Albumin < LLN

Severe – Total Bilirubin > ULN and Albumin < LLN, where baseline values of total bilirubin and albumin are taken to be the average of all pre-dose measurements.

- The CLIF-C ADs is derived using the following formula:

CLIF-C ADs = 10 × 0.03 × [Age (years)] + 0.66 × ln[creatinine (mg/dL)] +

1.71 × ln[INR] + 0.88 × ln[WBC (10^9 cells/L)] – 0.05 × [sodium (mmol/L)] + 8.

Inverse Probability of Censoring Weighting

The hazard ratio will be estimated using Inverse Probability of Censoring Weighting (IPCW) techniques (Robins 1999). Baseline predictors of both censoring and clinical outcome will be identified in advance. Augmented IPCW estimators described by Robins may be implemented to explore the possible impact of unmeasured confounding. Baseline predictors of both censoring and clinical outcome for the IPCW analysis are:

- Treatment Group: OCA/Placebo

- Age (year) at Screening Visit

- Sex: Male/Female

- UDCA use at Screening Visit: Yes/No

- Liver biochemistry at Baseline (alkaline phosphatase (ALP) (U/L), alanine transaminase (ALT) (U/L), aspartate transaminase (AST) (U/L), total bilirubin (µmol/L), albumin (g/L), and platelets)

The post-baseline time-varying covariates are ALP, ALT, AST, total bilirubin, and AE of pruritus with severity of moderate and severe.

To derive the IPCW weights, the patient's follow-up time up until the time of censoring or event occurring will be partitioned into about 10 intervals. The probability of remaining uncensored at the end of each time interval adjusted for baseline variables, and post-baseline time-varying covariates will be estimated using a pooled logistic regression model. To avoid possible extreme values when taking the inverse of the estimated probabilities from the pooled logistic model with both baseline and time-varying post-baseline covariates, the inverse of these probabilities will be stabilized by multiplying the probability of remaining uncensored conditional only on baseline covariates. Once the IPCW weights are derived, the hazard ratio of treatment effect and corresponding 95% CI will be estimated using a weighted Cox regression model with Sandwich estimator to obtain the robust estimate of the variance-covariance matrix of the parameter estimates.

Multiple Imputation (Missing Not At Random) Description

The event rate during off-treatment period (after the treatment discontinuation and before the study close-out) is calculated. The survival time during off-treatment period is assumed to follow exponential distribution of calculated event rate. The survival time during off-treatment

period is added to the survival time during on-treatment period for those subjects who require imputation (ie, early terminators). If the total duration is longer than the period from randomization to study close-out, the survival time will be imputed to the study close-out as a censored observation. Otherwise, the survival time will be imputed to the date of event. Multiple imputation is used to impute the survival time by 5000 iterations.

Procedures:

Step	Description
1	From the ADSL dataset restricted to ITT population (ITTFL='Y'), extract patient identifiers (USUBJID), randomization date (RANDDT), randomized treatment (TRT01PN), death date (DTHDT), treatment (TRT01PN) start (TRTSDT) and stop (TRTEDT) date for treatment, study termination date (WTHDRDT), date of last contact (LASTDT), follow-up after treatment discontinued (FUFL), and stratification factors. Sort dataset by USUBJID.
2	From the efficacy analysis dataset (ADTTE), select the primary endpoint event. Select patient identifiers (USUBJID), randomized treatment (TRT01PN), the censoring indicator (CNSR), follow-up length (ADY), and last contact date (ADT). Sort dataset by USUBJID.
3	Identify patients that discontinued study drug (IPDISC='YES') from the ADDS dataset. Retain USUBJID, analysis start date (ASTDT), and IPDISC variables. Sort dataset by USUBJID.
4	Merge ADSL, ADTTE, and ADDS data together to create a dataset for analysis named EFF1. • Create a variable for start of close out period (e.g., 12 Feb 2022). • Assess if last contact date (ADT) is non-missing and less than end of study date o If ADT non-missing and less than study end and patient was censored in the primary analysis then create a variable called COMPLETER and assign a value of 0 (study termination). o If ADT non-missing and greater than study end date, or the ADT is greater than the closeout date, or if non-censored in primary analysis then create a variable called COMPLETER and assign a value of 1 (primary endpoint event/completed study). • Assess if patient discontinued study medication (IPDISC='YES'), or was randomized and not treated, and evaluate this based on the censoring status in primary analysis and the COMPLETER status described above. • If IPDISC='YES' and patient was censored in primary analysis (CNSR=1) and did not complete the study (COMPLETER=0) then create a variable indicating the follow-up data will be imputed for the patient (IMPUTE=1). - Create a variable called EVENT and set to missing to hold the censoring indicator used in analysis following multiple imputation. - Create a variable called T2EVENT and set to missing to hold the follow-up time used in analysis following multiple imputation. - Create a variable called T0 to hold the follow-up time from primary analysis (T0=ADY)

Step	Description
	- Create a variable called T1 to hold the follow-up time from primary analysis (if DTHDT>. then T1=DTHDT – ADT + 1, otherwise T1= 12 Aug 2022 – ADT + 1).
	• If patient had a primary endpoint event in the primary analysis (CNSR=0), or completed the trial (CNSR=1 and COMPLETER=1), or patient did not discontinue drug prior to start of close out period (IPDISC ^= 'YES') and COMPLETER=1) then create a variable indicating the follow-up data will be not imputed for the patient (IMPUTE=0).
	- Create a variable called EVENT and set equal to 1 (EVENT=1) if patient experienced a primary endpoint event in primary analysis (CNSR=0), and 0 otherwise (EVENT=0).
	- Create a variable called T2EVENT and set to equal to the follow-up time from primary analysis (T2EVENT=AVAL).
	Sort dataset by treatment (TRT01PN) and randomization strata
5	For patients with IPDISC='Y', and remained in follow-up after discontinuation (ADT>TRTEDT and (DTHDT=. or DTHDT> ADT)), or for patients randomized and not treated that remained in follow-up IPDISC='Y' and (ADT>. and TRTSDT=. and (DTHDT=. or DTHDT> ADT)), sum the number of primary endpoint events (EVENT) by treatment group and randomization strata. Output dataset named EFF_COUNT.
6	*For those same patients from Step 5* sum the person-years of follow-up from discontinuation (TRTEDT) or from RANDDT (when TRTSDT=.) to the earliest of a primary endpoint event, last visit, death, or start of close out period (ADT). Output dataset named EFF_EXP.
7	Create a dataset called EFF_RATE merging EFF with EFF_COUNT and EFF_EXP *by treatment group and primary analysis strata*. Derive hazard rates for primary endpoint event for each patient in a new variable named RATE. Set RATE=365.25*(primary endpoint event sum/follow-up time sum).
8	Create a new dataset called IMPUTE from EFF_RATE restricted to patients that required follow-up imputation (IMPUTE=1). Derive 5,000 follow-up times for each patient for imputation in a variable named T2 as T2 is the randomly sampled from the exponential distribution given the event rate. When generating the exponential random variate specify a SEED equal to 1234 prior to derivation of T2. Assign the sequence of the T2 generated to a variable named ITERATION.
9	Create a dataset called IMPUTE2 from the IMPUTE dataset. For each of the 5,000 observations created in Step 8, derive variables EVENT=1 and T2EVENT=T0+T2 if T2 is less than the time from study termination (LASTDT) to the earlier of the start of close out period or (post-termination) death (T2< DTHDT - LASTDT if DTHDT>, or T2<Aug 12, 2022 – LASTDT). Otherwise, set T2=0 and T2EVENT equal to the time from randomization to the earlier of the start of close out period, or post-termination) death (whichever occurs first). Here T2EVENT= DTHDT - RANDDT+1 if DTHDT> LASTDT, otherwise T2EVENT=Aug 12, 2022 – RANDDT+1
10	Create 5,000 (virtual) datasets called IMPUTE3 by concatenating IMPUTE2 (for each iteration) with EFF with restriction to subjects that did not terminate the study (separately). For each of the 5,000 datasets perform a primary analysis (fit a stratified Cox model or KM method). Save parameter estimates for each iteration, and save the iteration variable.

Step	Description
11	Concatenate output from Step 10 and calculate p-value. Combine using Rubin's rule via PROC MIANALYZE.
12	Table results from Step 11 and stop.

Intercept

CLINICAL PHARMACOLOGY ANALYSIS PLAN
End of Study Noncompartmental Analyses for Study 747-302

Version 1: 12 April 2022

Study Protocol Number: **747-302**

Study Title: **A Phase 4, Double-Blind, Randomized, Placebo-Controlled, Multicenter Study Evaluating the Effect of Obeticholic Acid on Clinical Outcomes in Subjects with Primary Biliary Cholangitis**

THE COBALT STUDY

Clinical Outcomes with OBeticholic Acid in Liver Treatment (COBALT)

PPD

4/12/2022

PPD

Date

Director, Clinical Pharmacology
Intercept Pharmaceuticals, Inc.

This document has been reviewed and approved by:

PPD

4/12/2022

PPD   PhD   Date
PPD   Clinical Research
Intercept Pharmaceuticals, Inc.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Term
AE	adverse event
ALP	alkaline phosphatase
ALQ	above limit of quantification
AUC_{0-6h}	area under the concentration versus time curve from time 0 to 6 hours
AUC_{0-24h}	area under the concentration versus time curve from time 0 to 24 hours
BLQ	below the limit of quantification
C4	7α-hydroxy-4-cholesten-3-one
CI	confidence interval
CPAP	Clinical Pharmacology Analysis Plan
CP	Child-Pugh
CP-A	Child-Pugh Class A cirrhosis
CP-B	Child-Pugh Class B cirrhosis
CP-C	Child-Pugh Class C cirrhosis
CSPH	clinically significant portal hypertension
C_{trough}	the concentration immediately prior to administration of the next dose
CV	coefficient of variation
glyco-OCA	glycine conjugate of OCA
HVPG	hepatic venous pressure gradient
LC/MS/MS	liquid chromatography and tandem mass spectrometry method
LS	least-squares
MELD	model for end-stage liver disease
NCA	noncompartmental analysis
OCA	obeticholic acid
PBC	primary biliary cholangitis
PD	pharmacodynamic
PK	pharmacokinetic
QC	quality control
SD	standard deviation
SMQ	Standardized MedDRA Query
tauro-OCA	taurine conjugate of OCA
TIPS	transjugular intrahepatic portosystemic shunt
T_{max}	time to achieve peak (maximum) plasma concentration
UDCA	ursodeoxycholic acid
ULOQ	upper limit of quantification
ULN	upper limit of normal

1.   STUDY TITLE

A Multicenter, Open Label, A Phase 4, Double-Blind, Randomized, Placebo-Controlled, Multicenter Study Evaluating the Effect of Obeticholic Acid on Clinical Outcomes in Subjects with Primary Biliary Cholangitis

The COBALT Study: Clinical Outcomes with Obeticholic Acid in Liver Treatment (COBALT)

Study Protocol Number: 747-302

2.   SCOPE OF ANALYSIS PLAN

This Clinical Pharmacology Analysis Plan (CPAP) defines the end of study noncompartmental analysis (NCA) of pharmacokinetics (PK) of Study 747-302.

Background and Scope

Obeticholic acid (OCA) was approved under the tradename Ocaliva® on 27 May 2016 for the treatment of primary biliary cholangitis (PBC) in combination with ursodeoxycholic acid (UDCA) in adults with inadequate response to UDCA, or as monotherapy in adults unable to tolerate UDCA. The indication was approved under the Accelerated Approval regulations based on a reduction in alkaline phosphatase (ALP). Study 747-302 (COBALT) is a postmarketing requirement study conducted to evaluate the effect of OCA on clinical outcomes in subjects with PBC.

Predose/trough plasma samples were collected on Day 1, and at Month 3, 6, 9, and 12; during the follow-up period predose plasma samples were taken annually at 12-month intervals until the end of the study. At Month 9 subjects were given the opportunity to opt into a PK sub study and have plasma samples collected prior to dosing and at 0.5, 0.75, 1, 1.5, 2, 2.5, 3, 4, 5, and 6 hours postdose.

This CPAP describes the evaluation of the data, execution of NCA, and the reporting of pharmacokinetic results. For various reasons, the actual analysis may deviate from the guidance in this document. Deviations from this CPAP will be documented in the clinical study report.

Biomarker (C4, ALP, bile acids, etc.) values will be analyzed as specified in the statistical analysis plan. Further analysis of any PK/ pharmacodynamic (PD) relationships may be explored as the data suggest.

3.   STUDY OBJECTIVES

This section presents the study objectives from the protocol to provide context for the scope of this CPAP, as described in Section 2.

3.1.   Primary Objectives

To assess the effect of OCA compared to placebo, in conjunction with established local standard of care, on clinical outcomes in subjects with PBC as measured by time to first occurrence of any of adjudicated events as specified in the Study 747-302 final SAP.

3.2.   Secondary Objectives

In addition, the SAP specified secondary objectives which includes:

- To characterize the pharmacokinetics of OCA and its conjugates in a subset of subjects.

4.   STUDY DESIGN

This study started pre-approval as a Phase 3 and continued post-approval as a Phase 4, double-blind, randomized, placebo-controlled, multicenter study to evaluate the effect of OCA on clinical outcomes in subjects with PBC. Subjects were screened twice during a 1 to 8-week Screening period prior to entering the study to allow for the collection of repeat serum chemistry samples (at least 2 weeks apart) to confirm pretreatment ALP and total bilirubin values.

Investigational product was taken orally, once daily for the majority of subjects; dose and frequency were modified for subjects with cirrhosis (including subjects progressing to cirrhosis during the study) and classified as Child-Pugh B or C. The randomization was stratified by ursodeoxycholic acid (UDCA) treatment (yes/ no) and baseline bilirubin categories (> upper limit of normal [ULN]/ ≤ ULN).

Figure 1:   Study Diagram:
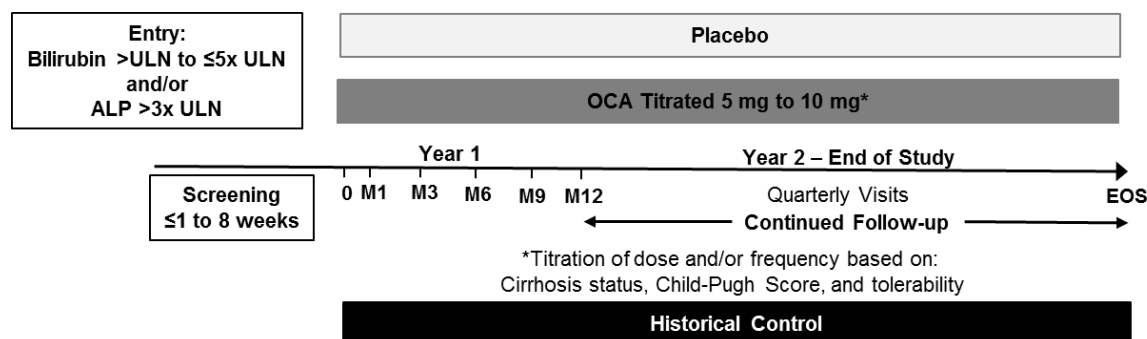
EOS = end of study; ULN = upper limit of normal

Initial dose titration of investigational product should occur at the Month 3 visit, or any study visit thereafter for subjects on all dosing regimens, based on tolerability and biochemical response. Subsequent dose titration(s) for subjects classified as Child-Pugh B or Child-Pugh C and following a modified dosing schedule may occur no earlier than 6 weeks after the previous dose titration.

Dosing frequency was determined by the presence or absence of cirrhosis and, if cirrhosis was present, by CP Score as described below:

- Noncirrhotic subjects or subjects classified as CP-A at screening will receive 5 mg OCA or matching placebo once daily for 3 months. Subjects should titrate to a maximum dose of 10 mg OCA once daily (or matching placebo) at the Month 3 visit or at any study visit following the Month 3 visit based on tolerability and biochemical response of the product.

- For those subjects that up-titrate to 10 mg, dosing may be decreased to 5 mg at any time during the study as considered clinically appropriate (e.g., based on tolerability). Subjects may be titrated back to a maximum dose of 10 mg once daily based on tolerability and clinical judgment of the Investigator.
- Subjects who are cirrhotic and classified as CP-B or CP-C will initiate a modified treatment regimen with 5 mg OCA or matching placebo once weekly for at least 3 months. Subjects classified as CP-B should eventually titrate to a maximum dose of 5 mg OCA or matching placebo once daily, based on tolerability and biochemical response. Subjects classified as CP-C should titrate to a maximum dose and frequency of 10 mg OCA or matching placebo twice weekly, based on tolerability and biochemical response.

Table 1:   Planned Dosing Regimen by Cirrhosis and Child-Pugh Score

	Planned Dosing Regimen		
	Standard	**Modified**	
	Noncirrhotic/ Child-Pugh A	**Child-Pugh B**	**Child-Pugh C**
Starting Dose [a] (Day 0)	5 mg daily	5 mg once weekly	5 mg once weekly
Titration 1 [b] (≥Month 3)	10 mg daily	5 mg twice weekly	5 mg twice weekly
Titration 2 [b] (≥6 weeks after Titration 1)	NA	10 mg twice weekly	10 mg twice weekly
Titration 3 [b] (≥6 weeks after Titration 2)	NA	5 mg daily	NA

[a] Starting dose based on subject's cirrhosis status and Child-Pugh Score at Screening.
[b] Planned titration regimen is shown; however, the titration of dose and/or frequency is dependent on subject tolerability, biochemical response, and/or changes in cirrhosis status or Child-Pugh Score at any time during the study

5.   SUBJECTS

The study aimed to enroll approximately 428 subjects. At study end, approximately 334 subjects were enrolled, and approximately 40 consented to the serial PK substudy.

6.   PK BACKGROUND

Subjects enrolled in the study had trough measurement s of OCA and the glyco- and tauro-conjugates determined at 3-month intervals for the first year of the study, and annually thereafter until the end of study. A substudy at 9 months was initiated to determine a plasma concentration/time curve from predose to 6 hours, and to allow imputation of a 24-hour curve.

7.   PK SAMPLING

Predose/trough plasma samples were collected on Day 1, and at Month 3, 6, 9, and 12; during the follow-up period predose plasma samples will be taken annually into the follow-up period and at the end of the study.

At Month 9/Week 36 a voluntary PK substudy was conducted at select sites. Subjects had plasma samples collected prior to dosing and at 0.5, 0.75, 1, 1.5, 2, 2.5, 3, 4, 5, and 6 hours postdose. Plasma concentrations of unconjugated OCA, glyco-OCA, and tauro-OCA will be measured using a validated liquid chromatography and tandem mass spectrometry method (LC/MS/MS). Drug concentrations will be reported from the laboratory in units of ng/mL.

In addition, a trough PK sample for the assessment of systemic concentrations of OCA and its major metabolites should be obtained in any patient who develops an adverse event (AE) that is indicative of or consistent with hepatic injury or decompensation.

Subjects in Study 747-302 were randomized to treatment with either OCA or matching placebo in a 1:1 ratio.

8.   DATA HANDLING

8.1.   Missing Values

The 6-hour PK profile has been optimized for characterization of the early part of the PK profile. A 24-hour profile can be generated by using the predose sample value as the 24-hour timepoint and thus estimating the AUC_{0-24h}. Missing values prior to the first dose will be set to zero; any other missing value will be set to missing and no imputation will be conducted.

8.2.   Results Below Limit of Quantification (BLQ)

BLQ values will be imputed to zero for all timepoints prior to the first quantifiable value. BLQ values occurring between 2 quantifiable values will be set to missing. Those BQL values that occur after the last quantifiable value will be set to ½ the lower limit of quantification (LLOQ). Other BLQ approaches may be used with rationale and documentation.

8.3.   Results Above Limit of Quantification (ALQ)

Typically, PK and PD samples with concentrations greater than the upper limit of quantification (ULOQ) are diluted into the range of the bioanalytical assay. However, if insufficient sample volume prevents repeat testing of the ALQ sample, the sample will be imputed to the AQL of the assay.

8.4.   Handling Anomalous Values

Drug concentration values that appear anomalous may be queried to the bioanalytical laboratory to see if there is a procedural explanation for the incongruous value(s). No repeat testing will be performed on samples for which there is no documented sample collection or bioanalytical error. Handling of anomalous values will be at the discretion of the Clinical Pharmacology Scientist. Final treatment of the anomalous values must be documented in the clinical study report.

8.5.　Calculation of Total OCA

Total OCA will be calculated as the molar sum of unconjugated OCA, glyco-OCA, and tauro-OCA, converted back to ng/mL-equivalents of the parent unconjugated OCA.

1. Convert units for each component of total OCA from ng/mL to μM using the molecular weight and formula:

   Unconjugated OCA (μM) = unconjugated OCA (ng/mL) / 420.6 g/mol

   Glyco-OCA (μM) = glyco-OCA (ng/mL) / 477.7 g/mol

   Tauro-OCA (μM) = tauro-OCA (ng/mL) / 527.8 g/mol

2. Sum molar-based components of total OCA:

   Total OCA (μM) = unconjugated OCA (μM) + glyco-OCA (μM) + tauro-OCA(μM)

3. Convert total OCA from μM to ng/mL-equivalents:

   Total OCA (ng/mL) = total OCA (μM) * 420.6 g/mol

Note that if there is not a valid value (BLQ imputations are valid) for one or more of the components of total OCA (unconjugated OCA, glyco-OCA, or tauro-OCA) then total OCA will not be calculated.

8.6.　Number Precision

Drug concentration values from the bioanalytical laboratory will not be rounded prior to use in NCA. Total OCA AUC_{0-6h} and AUC_{0-24h} values will be reported in the outputs from WinNonlin (Section 11) to the default precision dictated by the WinNonlin software.

8.7.　Actual versus Nominal Times

Actual collection times will be used for NCA. In cases where nominal times are used, a rationale will be documented.

9.　SOFTWARE USED

NCA-ready datasets may be produced using SAS Version 9.4 or higher. NCA will be performed using Phoenix WinNonlin Version 8 or higher.

10.　SUBJECT POPULATION TO BE ANALYZED

The process described herein will be executed for the PK populations as described in the protocol.

- The PK population will include all OCA subjects who have at least one confirmed fasted analyzable sample. Subjects must have been fasting for approximately 8 hours prior to the visit and must not have any major protocol deviations that potentially affect exposure levels. The PK population will be used for OCA PK analyses.

11.   PROCESS FOR ANALYSES

The end of study NCA for Study 747-302 will follow the steps described below.

- An analysis-ready dataset is created by merging the dosing and PK sampling data with the drug concentration data for unconjugated OCA, glyco-OCA, tauro-OCA, and total OCA
  - Units for concentrations will be ng/mL and units for time will be hours
  - Actual time since last dose will be calculated as the time variable. While PK parameters will be calculated for PK profiles with deviations from nominal times, gross deviations may render PK parameters incompatible for inclusion in summary statistics. In such a case, the parameters in question would be included in listings (and flagged) but excluded from summary tables and figures.
  - BLQ values will be imputed as per Section 8.2.
  - Total OCA will be calculated as per Section 8.5. In general, Categorical variables will be summarized by counts and by percentage of subjects in corresponding categories.

Evaluation of the data should include:

- The drug concentrations and PK parameters will be available for:
  - Unconjugated OCA
  - Glyco-OCA
  - Tauro-OCA
  - Total OCA (molar sum of OCA, glyco-OCA, tauro-OCA)
- Examples of potential reasons for exclusion of data from PK summary analyses include, but are not limited to:
  - Subject not fasted for 8 hours prior to PK visit
  - Subjects not adhering to standardized meal timing (or missed meal) during PK profile collection
  - Non-standard dosing (e.g., drug holiday, altered dose level, altered dose frequency) used prior to PK visit

  Data excluded based on this review will be included in listing outputs (and flagged with footnote) but will be excluded from summary tables and figures.

- Drug concentration data in ng/mL units will be summarized by analyte (unconjugated OCA, glyco-OCA, tauro-OCA, and total OCA), baseline CP category (or noncirrhotic), dosing regimen, study visit, and timepoint using the following descriptive statistics: n, number and % BLQ, arithmetic mean, SD, CV%, minimum, median, maximum, interquartile range, geometric mean, and geometric CV%.
- Trough PK samples for the assessment of plasma concentrations of OCA and its conjugates, obtained in any subject who develops an AE that is indicative of or

consistent with hepatic injury or decompensation, will be summarized separately from scheduled PK collections and stratified by baseline CP category (or noncirrhotic) and dosing regimen.

- Actual collection date and time, time since reference dose, time deviation between actual and nominal collection times, flags for exclusion from summaries, and stratification variables should be included.

- PK concentration summary tables presenting descriptive statistics over time by analyte and including stratification by any relevant factors.

PK parameters will be calculated and examined using the execution of NCA (Section 11.2). The data will be summarized as follows:

- PK parameters will be calculated in subjects at Month 9 for whom a PK profile can be produced. Parameters to be calculated are found in Table 2.

- Individual PK parameters for unconjugated OCA, glyco-OCA, tauro-OCA, and total OCA will be listed by subject and include information for baseline CP category, dosing regimen, study visit, and timepoint.

- PK parameters will be summarized by analyte (unconjugated OCA, glyco-OCA, tauro-OCA, or total OCA, PK parameter, baseline CP Category (noncirrhotic, CP-A, CP-B, or CP-C), baseline MELD category ($>/\leq 8$ and $>/\leq 11$), dosing regimen (dose strength/frequency), and study visit. For the "All Visits" category, if an individual subject has data from multiple visits on a particular dosing regimen, the mean of the values across those visits will be used so that each subject is only represented once. The summary statistics to be used for each PK parameter are described in Table 3. Reporting precision is generally to 3 significant figures except for the following: time to achieve peak (maximum) plasma concentration (T_{max}) is reported to same precision as nominal time, SD is reported to same precision as mean and median, and CV and geometric CV are reported to the same precision as mean and geometric mean, respectively.

- The following figures will be produced for mean concentration data for all analytes:
  - Mean (+SD) 6-hour and 24-hour PK profiles for unconjugated OCA, glyco-OCA, tauro-OCA, and total OCA will be provided as follows:
  - Nominal hours on continuous x-axis
  - Mean (+SD) plasma concentration (ng/mL) on linear y-axis
  - Unconjugated OCA, glyco-OCA, tauro-OCA, and total OCA are presented on same figure (5 lines)
  - Outputs should present one visit per page which will involve a matrixed panel of figures to stratify for CP category (or noncirrhotic) and dosing/regimen at each visit.

- Boxplots of the relevant AUCs (AUC_{0-6h} and AUC_{0-24h}) will be provided for each PK study visit, stratified by baseline CP category (or noncirrhotic) and the dosing regimen used at the time of the PK visit. Separate figures will be produced for

unconjugated OCA, glyco-OCA, tauro-OCA, and total OCA. Boxplots will also include the individual AUC values from each subject.

- Boxplots of C_{trough} will also be generated. Separate figures will be generated for unconjugated OCA, glyco-OCA, tauro-OCA, and total OCA at each PK study visit (Day 0, Month 3, Month 6, Month 9, Month 12, and annual visits thereafter). Stratification within each figure will be by CP category (noncirrhotic and CP-A subjects only) and the dosing regimen used at the time of the PK visit. Boxplots will also include the individual C_{trough} values from each subject.

- Statistical comparisons (e.g., Proc Mixed) of the ln-transformed PK parameters from Study 747-302 will compare exposures from the following groups: noncirrhotic 5 mg OCA daily, CP-A 5 mg OCA daily, noncirrhotic 10 mg OCA daily, and CP-A 10 mg OCA daily. Separate analyses will be performed for each PK parameter/Visit to include total OCA AUC_{0-6h} from Month 9 and total OCA C_{trough} from Month 3, Month 6, Month 12, or subsequent annual visits (provided sufficient data are available). The geometric least-squares (LS) means, geometric LS mean ratios (test/reference), and associated 90CIs will be presented and will be exponentiated to the original scale. The reference group presented will alternately be the 5 mg OCA daily noncirrhotic group or the 10 mg OCA daily noncirrhotic group. Additional analyses may be generated if deemed appropriate.

- PK parameter summaries may be re-analyzed as a function of the key subgroups following the preliminary PK analysis.

- Depending on the information available for biomarkers, PK/PD analyses may be performed if deemed appropriate. The description of these analyses will be provided on an ad hoc basis.

In addition, trough PK samples for assessment of drug concentrations of OCA and its conjugates, obtained in any patient who develops an AE that is indicative of or consistent with hepatic injury or decompensation, will be identified as such in the listings. The actual and nominal sampling times of PK sample collection will also be listed for each subject and will include the deviation in time from nominal, if applicable. All measured concentrations will be presented in original units as reported by the bioanalytical lab (i.e., ng/mL). Columns in the listing for unconjugated OCA, glyco-OCA, tauro-OCA, and total OCA will be displayed on the same page. Any individual concentrations excluded from summaries or analyses will be flagged and footnoted.

If the values for one or more of the components of total OCA are missing, then total OCA will not be calculated and will be set to missing. If the value for one or more of the components of total OCA are BLQ then appropriate BLQ imputation will be performed prior to the calculation of total OCA. Total OCA will be calculated in units of "unconjugated OCA ng/mL equivalents" but for the sake of brevity, total OCA results will be reported in units of "ng/mL."

Table 2:   **PK Parameters to be Calculated for Study 747-302**

PK Parameter	Units	Definition
AUC_{0-6h}	ng*h/mL	Area under the concentration versus time curve from zero time to 6 hours. The linear/linear trapezoidal rule should be used for estimation of AUC. At least 4 quantifiable concentration-time values must be available to compute AUC_{0-6h}
AUC_{0-24h}	ng*h/mL	Area under the concentration versus time curve from zero time to 24 hours. The predose sample for each profile is to be used as the 24-hour timepoint The linear/linear trapezoidal rule should be used for estimation of AUC. At least 4 quantifiable concentration-time values must be available to compute AUC_{0-24h}
C_{max}	ng/mL	Maximum observed concentration. If all observations are BLQ, C_{max} will be reported as zero
C_{trough}	ng/mL	Pre-dose concentration
T_{max}	H	First time at which C_{max} is observed. If all observations are BLQ, T_{max} will be reported as not determined (ND)
MR_{AUC}	unitless (ratio)	Molar metabolite to parent ratio for AUC_{0-6h}
MR_{Cmax}	unitless (ratio)	Molar metabolite to parent ratio for C_{max}

Table 3:   **PK Parameter Summary Statistics**

Variable	Summarized with:
AUC_{0-6h}, AUC_{0-24h}, C_{trough}, C_{max}, MR_{AUC}, and MR_{Cmax}	n, arithmetic mean, SD, CV%, minimum, Q1 (25% percentile), median, Q3 (75% percentile), maximum, geometric mean and geometric CV%
T_{max}	n, minimum, Q1 (25% percentile), median, Q3 (75% percentile) and maximum

11.1.   Data Presentation Conventions

The precision of original measurements will be maintained in summaries, when possible. Means, medians, and SDs will be presented with an increased level of precision, where means and medians will be presented to 1 more decimal place than the raw data, and the SDs will be presented to 2 more decimal places than the raw data. In general, the number of decimal places should not exceed 3 decimal places, unless deemed appropriate. For tables where rounding is required, rounding will be done to the nearest round-off unit. For example, when rounding to the

nearest integer, values ≥XX.5 will be rounded up to XX + 1 (e.g., 97.5 will round up to 98), while values <XX.5 will be rounded down to XX (e.g., 97.4 will round down to 97).

Percentages based on frequency counts will be based on available data, and denominators will generally exclude missing values. Percentages based on frequency counts will be presented to 1 decimal place, and values less than 0.1% will be presented as "<0.1%." Values less than 100% but that round up to 100% will be presented as ">99.9%." Date variables will be formatted as YYYY-MM-DD for presentation. Time will be formatted in military time as HH:MM for presentation.

11.2.   Execution of NCA

Using the analysis-ready dataset, NCA will be performed according to the following:

- Review data on fed status prior to dosing and prior to 24-hour sampling

- Review data on dose (mg) and timing of meal during PK profile

- Exploratory data analysis will be performed by creating XY plots of concentration versus actual time, grouped by analyte (unconjugated OCA, glyco-OCA, tauro-OCA, and total OCA) to inspect the PK profiles

- NCA will be performed for each analyte with the following settings:
  - Linear trapezoidal/linear interpolation methodology
  - 'Disable Curve Stripping' setting to avoid calculation of an elimination slope
  - Use of actual time since last dose as time variable (hours)

- Calculate the AUC_{0-6}. Estimate AUC_{0-24h} by using the predose value and setting that value to 24 hours.

- Carry the study, subject, and visit variables into the analysis for output purposes

12.   FILE STORAGE

The work associated with this plan will be stored in the standardized folder structure on the SASShare drive at the following project folder location:

\Compounds\INT747\POPPK\ 747-302

The saved work will be organized in such a manner that it is easy to find, for any given subject, the Primary, QC, and final NCA results along with the associated signed QC form.

12.1.   Key Subgroups for Evaluation

To fulfill the scope of the analysis, the following subgroups have been identified for the evaluation based on known association of risk with clinical outcomes PBC or end-stage liver disease. Further PK assessment by these groups may be required, based on outcomes.

<u>Subgroups Defined by Baseline Disease Severity:</u>

- Cirrhotic vs Noncirrhotic

  - Cirrhosis as defined by baseline status on the cirrhosis case report form and screening/baseline CP Evaluation

- Noncirrhotic vs Child Pugh A vs Child Pugh B

  - Subjects who received daily dosing vs those who received the modified dosing regimen

- Noncirrhotic vs Child Pugh A5 vs Child Pugh A6 vs Child Pugh B7 vs Child Pugh B8 vs Child Pugh B9

- Baseline/Medical History of clinically significant portal hypertension (CSPH) vs Patients with No Baseline/Medical History of CSPH defined by any of the following:

  - Previous history of portal hypertension includes procedures for transjugular intrahepatic portosystemic shunt (TIPS), sclerotherapy, ligation, hepatic venous pressure gradient (HVPG) measurement >10 mmHG, paracentesis, thoracentesis (due to hepatic hydrotorax)

  - Baseline biochemical evidence of portal hypertension

  - Previous history or baseline observation of the terms defined in Table 4.

Table 4:　　Adverse Event Terms Associated Used to Define Presence or Absence of Clinically Significant Portal Hypertension (CSPH)

Presence of CSPH based on SMQ terms	Severity
Presence of collateral circles secondary to CSPH (SMQ Terms: Varices - Anorectal, Duodenal, Esophageal, Gastric, Intestinal, Splenic, Gallbladder, Peripancreatic – Portal Hypertensive Gastropathy and Enteropathy)	Any
GI Bleeding (SMQ Terms Gastric Variceal Hemorrhage, Oesophageal Variceal Hemorrhage, Anorectal variceal hemorrhage)	Any
Ascites (SMQ terms Ascites, Hepatic Hydrotorax, Oedema due to hepatic disease, Peritonitis Bacterial, bacterascites)	Any
Multiorgan complication: Hepatopulmonary syndrome, hepatorenal syndrome.	Any
Portopulmonary Hypertension	Any

- Baseline MELD >/≤ the following thresholds:
  - 8
  - 11
- Baseline Rotterdam Classification (mild vs moderate vs severe)
- Baseline GLOBE Score >/≤ the following thresholds:
  - Age Specific Thresholds
    - -0.52 for <45 years; 0.01 for ≥45 to <52 years; 0.60 for ≥52 to <58 years; 1.01 for ≥58 to <66 years; 1.69 for ≥66 years
  - 0.9
- Baseline Mayo Risk Score >/≤ the following thresholds:
  - 7.43
  - 8.50
  - 9.10
- Baseline Platelets >/≤ the following thresholds:
  - 150,000 10^9/L
  - 100,000 10^9/L
  - 50,000 10^9/L
- Baseline Albumin >/≤ the following thresholds:
  - 4 g/dL
  - 3.5 g/dL
  - 3 g/dL
- Baseline Bilirubin >/≤ the following thresholds:
  - ULN
  - 1.5xULN
  - 2xULN

Subgroups Defined by the Occurrence of Key On-Treatment Events:
- Evidence of CSPH
  - Yes vs No defined by any of the following criteria:
    - Including history of, baseline, or on treatment occurrence of terms described in Table 4
    - Including history of, baseline, or on treatment occurrence of procedure of TIPS, sclerotherapy, ligation, HVPG measurement >10 mmHG, paracentesis, thoracentesis (due to hepatic hydrotorax)
    - Baseline or on treatment biochemical evidence at baseline of portal hypertension
- On treatment occurrence of a composite negative outcome event (as defined by Table 4)
  - Yes vs No

In the situation where small cell count of certain subgroup occurs (cell count <5) in any of the subgroup analysis, it may not be deemed necessary for the subgroup analysis to be performed.

Intercept

Statistical Analysis Plan for Interim Analysis
for Protocol 747-302

A Phase 4, Double-Blind, Randomized, Placebo-Controlled, Multicenter Study Evaluating the Effect of Obeticholic Acid on Clinical Outcomes in Subjects with Primary Biliary Cirrhosis

OBETICHOLIC ACID (OCA)

Protocol Version and Date:	Version 6: 05 November 2019
Phase:	Phase 4
Methodology:	Double-Blind, Randomized, Placebo-Controlled Study
Sponsor:	Intercept Pharmaceuticals, Inc.
	4760 Eastgate Mall
	San Diego, CA 92121
	Tel: 858-652-6800
	Fax: 858-558-5961
Independent Statistician:	PPD Cytel
Study Statistician:	PPD PhD
	Sr. Manager, Biostatistics, Intercept Pharmaceuticals, Inc.
Analysis Plan Date:	20 August 2020
Analysis Plan Version:	Amendment 1.0

CONFIDENTIAL

APPROVAL

Upon review of this document, including table and listing shells, the undersigned approves the Statistical Analysis Plan. The analysis methods and data presentation are acceptable.

PPD

PPD PhD

Sept. 21, 2020

Date

Director, Biostatistics
Intercept Pharmaceuticals, Inc.

PPD

PPD PhD PPD

22 Sept. 2020

Date

Biostatistics and Data Management
Intercept Pharmaceuticals, Inc.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

Abbreviation	Term
ALP	alkaline phosphatase
CI	confidence interval
CPS	Child-Pugh Score
dL	deciliter
DMC	Data Monitoring Committee
DOB	date of birth
DOIC	date of informed consent
INR	international normalized ratio
ITT	Intent-to-Treat
KM	Kaplan-Meier
MELD	Model for End Stage Liver Disease
mg	milligram
OCA	obeticholic acid
PBC	primary biliary cirrhosis
SAP	statistical analysis plan
SD	standard deviation
SEM	standard error of the mean
UDCA	ursodeoxycholic acid
ULN	upper limit of normal

1.   INTRODUCTION

Study 747-302 is a double-blind, randomized, placebo-controlled, time-to-event study with two planned interim analyses and one final, end of study (EOS) analysis. According to the study protocol (version 6), two interim analyses were planned after accrual of approximately 50% (64 adjudicated events) and 75% (96 adjudicated events) of clinical outcome events, respectively; however, one interim analysis will be performed after the accrual of approximately 65% (83 adjudicated events) of clinical outcome events.

In addition to the interim efficacy analyses pre-specified in the protocol, the Sponsor also plans to include the following analyses at the time of the interim analysis: perform a sample size re-evaluation; evaluate the feasibility of the study due to treatment discontinuation, overall and specifically related to use of commercial OCA or baseline disease characteristics (ie, Child-Pugh score and Rotterdam Criteria). The rationale to perform these additional analyses are described in Section 6. Axio, an independent DMC vendor, will perform the interim analysis and the unblinded results of the interim analysis will be reviewed by the DMC. The DMC will make recommendations according to the criteria outlined in this interim statistical analysis plan (SAP). The DMC will also assess continued feasibility of the study. The Sponsor will remain blinded to the data generated at the time of the interim analysis.

This statistical analysis plan (SAP) outlines the specific methodology and the analyses for the interim analysis, which is focused on efficacy and the study feasibility assessment. The statistical analysis plan for the end-of-study analysis will be developed as a separate document. In addition to the analyses presented within this document, the Sponsor will evaluate the viability of including progression to cirrhosis in the primary efficacy endpoint in a blinded fashion. The method of evaluating progression to cirrhosis will be described in a separate document.

1.1.   Study Objectives

1.1.1.   Primary Objective

The primary objective of this study is to compare the effect of obeticholic acid (OCA) to placebo, in conjunction with established local standard of care, on clinical outcomes in subjects with primary biliary cholangitis (PBC) as measured by time to first occurrence of any of the following adjudicated events, derived as a composite event endpoint:

- Death (all-cause)
- Liver transplant
- Model of end stage liver disease (MELD) score ≥15
- Hospitalization (as defined by a stay of 24 hours or greater) for new onset or recurrence of:
  - Variceal bleed
  - Hepatic encephalopathy (as defined by a West Haven score of ≥2)
  - Spontaneous bacterial peritonitis (confirmed by diagnostic paracentesis)
- Uncontrolled ascites (diuretic resistant ascites requiring therapeutic paracentesis at a frequency of at least twice in a month)

1.2. Study Design

1.2.1. Synopsis of Study Design

This is a Phase 4, double-blind, randomized, placebo-controlled, multicenter study. Eligible subjects have a diagnosis of PBC with bilirubin levels of >ULN and ≤5x ULN and/or ALP >3x ULN. Subjects enrolled are at higher risk of liver-related clinical complications.

Approximately 428 subjects meeting all enrollment criteria will be recruited into the study over an approximate 4-year period, randomly allocated to treatment with either OCA or matching placebo in a 1:1 ratio. Randomization will be stratified by UDCA treatment (yes/no) and baseline bilirubin categories (>ULN/≤ULN). A minimum of 30% of subjects will have elevated bilirubin (>ULN) at Screening. In addition to the placebo control arm, multiple external control groups (concurrent and retrospective) will be used to compare OCA-treated patients to standard-of-care-treated patients.

Subjects will be dosed according to their cirrhosis status and Child-Pugh Score (CPS), receiving a maximum 10 mg OCA or matching placebo once daily, based on tolerability and biochemical response.

Subjects are expected to be followed for a minimum of approximately 6 years. The study will continue until approximately 127 adjudicated primary endpoint events have been accrued in unique subjects, or until the Sponsor (eg, based on a recommendation from the Data Monitoring Committee; DMC) terminates the study. One interim analysis is planned, after the accrual of ~65% (83 adjudicated events) of clinical outcome events.

This study will be conducted in a double-blind, placebo-controlled manner. The randomization will be stratified by UDCA treatment (yes/no) and Baseline total bilirubin categories (>ULN/≤ULN) where the normal range is specified by the central laboratory.

Schedules of study procedures are outlined in Table 1 (Screening to Month 12) and Table 2 (Year 2 to End-of-Study Endpoints) of the study protocol.

The protocol provides further study details.

1.2.2. Primary Efficacy Endpoint

The primary efficacy endpoint will be the time to first occurrence of one of the following post-randomization:

- Death (all-cause)
- Liver transplant
- Model of end stage liver disease (MELD) score ≥15
- Hospitalization (as defined by a stay of 24 hours or greater) for new onset or recurrence of:
    - Variceal bleed
    - Hepatic encephalopathy (as defined by a West Haven score of ≥2)
    - Spontaneous bacterial peritonitis (confirmed by diagnostic paracentesis)

- Uncontrolled ascites (diuretic resistant ascites requiring therapeutic paracentesis at a frequency of at least twice in a month)

If a subject experiences more than one of the outcomes above, the primary efficacy analysis will include only the outcome that occurs first.

2.   SAMPLE SIZE JUSTIFICATION

2.1.   Sample Size

The following assumptions were used in sample size calculations for the study:

- Exponential survival curves, placebo survival estimate of 0.6 at 8 years with a hazard ratio of 0.60 comparing between placebo and OCA treatments, and total study duration of 10 years, allowing for 4 years of accrual and 6 years of follow up of the last subject enrolled.

- Subjects will be randomized in a 1:1 ratio to placebo or OCA.

- The 2 treatment groups will be compared using a 2-sided log rank test at the 5% level of significance.

- Two interim analyses and one final analysis are planned, with interim analyses occurring after the accrual of 50% and 75% of clinical outcome events, respectively.

Based on the randomization ratio, significance level, and assumed hazard ratio, a total of 127 events (both groups combined) will provide 80% power to demonstrate a statistically significant difference between OCA and placebo on time to liver-related outcomes, including all-cause mortality.

Based on the remaining assumptions stated above, approximately 428 subjects will need to be enrolled to attain 127 events.

3.   ANALYSIS POPULATIONS

The following subject populations will be evaluated and used for presentation and analysis of the data:

- Randomized Population will include all randomized subjects.

- The Intent-to-Treat (ITT) Population will include all randomized subjects who receive any amount of investigational product (OCA or placebo). Treatment assignment will be based on the randomized treatment. For example, subjects who discontinue placebo treatment and start open-label or commercially marketed OCA will be analyzed as placebo subjects.

The primary endpoint analysis will be conducted in the ITT population.

4. EFFICACY ANALYSES

4.1. Primary Efficacy Analysis and Multiplicity Adjustment

The primary efficacy endpoint will undergo one interim analysis prior to final database lock and results will be reviewed by the independent DMC. A group sequential design using the O'Brien-Fleming type alpha-spending function (Lan & DeMets 1983) (Reboussin 2000) will be used to control the overall two-sided 0.05 significance level allocated to hepatic clinical outcomes (Table 1). The interim analysis will occur after accrual of approximately 65% (83 adjudicated events) of clinical outcome events (actual number of events based on observed data at interim analysis will be used for analysis) and the final analysis will occur after the accrual of 127 adjudicated clinical outcome events. To determine the timing for conducting these analyses, the number of pre-adjudicated endpoint events will be monitored and used for the projection.

Stopping criteria for efficacy at the planned timepoints based on group sequential boundaries are provided in Table 1 for the Type I error rate comprising the sum of prespecified alpha for clinical outcomes and carried forward alpha from the interim analysis. The exact alpha level and corresponding hazard ratio criterion at the time of the interim analysis will be calculated based on the actual observed number of clinical events.

The criteria for the DMC to recommend that the study stop for efficacy will be met at the interim analysis when the estimated hazard ratio is less than δ^*, where δ^* is the hazard ratio efficacy criterion calculated based on observed data at the interim analysis.

Table 1: **Stopping Criteria Based on O'Brien-Fleming Boundaries**

Overall Type 1 Error	Overall Power	Information Fraction	Number of Clinical Events	Incremental α Spent	Cumulative α Spent	Reject for Efficacy if Hazard Ratio
0.05	0.80	65%	83	0.011	0.011	≤0.564 ($\leq\delta^*$)
		End of Study (100%)	127	0.047	0.050	≤0.697

Note: Estimates for information fraction and number of clinical events at interim analysis are provided in Table 1 for illustration, and the hazard ratio efficacy criteria δ^* will be calculated based on observed data.

4.2. Time–to-Event Endpoints

For the time to event analyses, subjects who do not experience an event will be censored at the time of their last contact. Missing data will be assumed to be missing at random. Details of sensitivity analyses are described in Section 4.2.2.

4.2.1. Primary Efficacy Analysis

The primary efficacy analysis will compare OCA and Placebo treatment groups with respect to the primary efficacy endpoint using the ITT population. The analysis will only include adjudicated events. The number and percent of subjects censored and with events will be presented. Descriptive statistics will be presented for the time-to-event. The 2 treatment groups will be compared using a log rank test stratified by the randomization stratification factor, conducted at the 2-sided alpha level of significance described in Section 2.1. Kaplan-Meier

(KM) estimates of the distribution of the time-to-event will be tabulated and graphed by treatment group. The tabulation will include the KM estimate of the 25th, 50th (median), and 75th percentiles and corresponding 2-sided 95% CIs, where the percentiles can be estimated. KM tabulations will present time to event in days; KM graphs will present time to event in months. The proportionality of hazards will be assessed using Schoenfeld residuals. The hazard ratio and 95% CI will be estimated based on a Cox regression model stratified by randomization strata.

All subjects will be analyzed according to their randomized treatment assignment, regardless of actual treatment received.

Subjects without any documentation of events will be censored at the date of last contact. For subjects with an event, the earliest of the following event dates will be used:

- Death (all-cause)

- Liver transplant

- Model of end stage liver disease (MELD) score ≥15

- Hospitalization (as defined by a stay of 24 hours or greater) for new onset or recurrence of:

  - Variceal bleed

  - Hepatic encephalopathy (as defined by a West Haven score of ≥2)

  - Spontaneous bacterial peritonitis (confirmed by diagnostic paracentesis)

- Uncontrolled ascites (diuretic resistant ascites requiring therapeutic paracentesis at a frequency of at least twice in a month)

If a subject experiences more than one of these events, only the event that occurs first will be included in the primary analysis.

4.2.2.   Sensitivity Analyses

The following sensitivity analyses will be carried out on the primary endpoint. They are intended to assist the DMC in assessing potential bias in the study:

1. The hazard ratio will be estimated using Inverse Probability of Censoring Weighting (IPCW) techniques (Robins 1999). The DMC will use this analysis to assess the impact of censoring on the hazard ratio. Baseline predictors of both censoring and clinical outcome will be identified in advance. Augmented IPCW estimators described by Robins may be implemented to explore the possible impact of unmeasured confounding. Baseline predictors of both censoring and clinical outcome for the IPCW analysis are:

- Treatment Group: OCA/Placebo

- Age (year) at Screening Visit

- Sex: Male/Female

- UDCA use at Screening Visit: Yes/No

- Baseline Total Bilirubin (μmol/L)

- Baseline Alanine Transaminase (U/L)

- Baseline Aspartate Transaminase (U/L)

- Baseline Alkaline Phosphatase (U/L)

- Baseline Albumin (g/L)

- Baseline Platelets count

The post-baseline time-varying covariates are: ALP, ALT, AST, Total Bilirubin, and AE of pruritus with severity of moderate and severe.

To derive the IPCW weights, the patient's follow-up time up until the time of censoring or event occurring will be partitioned into about 10 intervals. The probability of remaining uncensored at the end of each time interval adjusted for baseline variables, and post-baseline time-varying covariates will be estimated using a pooled logistic regression model. To avoid possible extreme values when taking the inverse of the estimated probabilities from the pooled logistic model with both baseline and time-varying post-baseline covariates, the inverse of these probabilities will be stabilized by multiplying the probability of remaining uncensored conditional only on baseline covariates. Once the IPCW weights are derived, the HR of treatment effect and corresponding 95% CI will be estimated using a weighted Cox regression model with Sandwich estimator to obtain the robust estimate of the variance-covariance matrix of the parameter estimates.

2. The primary efficacy analysis will be repeated using only data accrued before subjects discontinued regularly scheduled study visits. The DMC will use this analysis to assess whether the time subjects spent off-treatment/off study-visits is affecting the hazard ratio.

3. The primary efficacy analysis will be repeated using only data accrued before subjects started commercial use of OCA. The DMC will use this analysis to assess whether the initiation of commercial use of OCA impedes the evaluation of the primary endpoint. The date a subject started commercial use of OCA will be taken to be the minimum of:

   a. The date of treatment discontinuation due to commercial OCA use
   b. The date of study visit discontinuation due to commercial OCA use
   c. The start date of OCA recorded as a concomitant medication

Kaplan-Meier (KM) estimates of the distribution of the time-to-event will not be graphed for any sensitivity analyses.

4.3. Additional Efficacy Analyses

4.3.1. Subgroup Analysis by Child-Pugh Classification

The primary efficacy analysis will be repeated within each of the baseline Child-Pugh Classification subgroups. Baseline Child Pugh classification is described in Section 7.4.

4.3.2.   Relationship between Time-to-Event Outcomes and Biochemical Response

Additional efficacy analyses of correlation of time-to-event outcomes and biochemical response will be run as supportive information for primary endpoint evaluation and in support of FDA PMR 3057-3 as described below.

A biochemical responder is defined as a subject who attains an ALP <1.67 times the upper limit of normal (ULN) and total bilirubin ≤ULN and an ALP decrease of ≥15% from Baseline at their 12-month visit. Subjects missing values are considered non-responders. The number and percent of subjects censored and with events and descriptive statistics for the time-to-event will be presented by biochemical response. KM estimates of the distribution of the time-to-event using the primary composite endpoint will be tabulated and graphed by treatment group and biochemical responder status, and by biochemical responder status only. KM tabulations will present time to event in days; KM graphs will present time to event in months.

Biochemical responders and non-responders will be compared using a log rank test stratified by treatment group, conducted at the 2-sided 0.05-level of significance. The proportionality of hazards will be assessed using Schoenfeld residuals and the hazard ratio and 95% CI will be estimated based on a Cox regression model stratified by treatment group.

The analysis will be repeated using the following alternative definitions of biochemical responder:

- ALP ≤ 2.0 x ULN at their 12-month visit
- Total bilirubin < 1 x ULN at their 12-month visit

For placebo patients, OCA-treated patients, and all ITT patients, additional KM estimates of the distribution of the time-to-event will be graphed by biochemical response for the combined ALP and bilirubin categories:

1. ALP ≤ 2.0 x ULN and Total bilirubin < 1 x ULN at their 12-month visit
2. ALP > 2.0 x ULN and Total bilirubin < 1 x ULN at their 12-month visit
3. ALP ≤ 2.0 x ULN and Total bilirubin ≥ 1 x ULN at their 12-month visit
4. ALP > 2.0 x ULN and Total bilirubin ≥ 1 x ULN at their 12-month visit

4.3.3.   Progression to Cirrhosis

The Sponsor will evaluate the viability of including progression to cirrhosis in the primary efficacy endpoint in a blinded fashion. The method of evaluating progression to cirrhosis will be described in a separate document.

5.   ANALYSES TO EVALUATE THE FEASIBILITY OF CONTINUING THE STUDY

5.1.   Study Discontinuation

Subject discontinuation, which will be tabulated by treatment group and overall, will include the number randomized, the number in ITT population for analysis, the number who withdrew from

investigational product and reason(s) for withdrawal, the number who started commercial use of OCA, the number who withdrew from study visits prior to completing the study and reason(s) for withdrawal, the number of subjects who were contacted for follow-up, the number who withdrew from the study, the number of subjects who withdrew from the study prior to a potential endpoint event, and the number of subjects who withdrew from the study prior to a positively adjudicated endpoint event. Subjects are considered to have withdrawn from the study if they discontinued study visits and did not consent to follow-up contact or medical record review or if they withdrew consent for follow-up contact or medical record review. Subjects are considered to have started commercial OCA if they discontinued treatment or study visits due to initiating commercial OCALIVA, or if OBETICHOLIC ACID is recorded as a concomitant medication. Percentages of the number treated in total within each category will be provided.

Subject discontinuation summaries by treatment group and overall will be repeated by baseline Child Pugh classification, baseline Rotterdam score, and by geography (region and country). Baseline Child Pugh classification and baseline Rotterdam scoring are described in Section 7.4.

Subject discontinuation information will be used by the DMC, among with other factors, to assess feasibility of continued conduct of the study as designed.

5.2.   Sample Size Recalculation

A sample size recalculation is included in order to re-assess the assumptions for the sample size and power evaluation. The original study sample size calculation assumed a 10-year study (4 years enrollment, 6 years follow-up) and 80% power to demonstrate a statistically significant difference between OCA and placebo on time to liver-related outcomes. The sample size recalculation described here allows for a 12-year study (10 years enrollment, 2 years follow-up) and 70% power to detect a difference between OCA and placebo on time to liver-related outcomes.

If the criteria for the DMC to recommend that the study stop for efficacy, ie, the estimated hazard ratio is less than δ^*, where δ^* is the hazard ratio efficacy criterion at the interim analysis, are not met at the interim analyses, the conditional power for rejecting the null hypothesis in favor of the alternative (ie, the probability of success at end of trial given the interim results) will be computed. If the conditional power is >70% at the interim analysis, the DMC will recommend that the study continue as is (ie, the study will not stop for efficacy nor will the sample size be increased). If the conditional power is less than 70% and the estimated hazard ratio is within the "promising zone" of δ^* and 0.85, the number of clinical endpoint events will be re-estimated using the methods described in Cui, Hung and Wang 1999 to achieve a conditional power of at least 70% for the end-of-study primary endpoint analysis. If the re-computed required number of events is less than 127, the original required number of events, 127, will be used. The total number of subjects to be enrolled in the study will be recalculated based on the new number of required endpoint events, observed placebo hazard rate at data cut-off of the interim analysis, assuming a total study duration of 12 years allowing for 10 years accrual and 2 years of follow up. The placebo hazard estimate will be calculated by fitting an exponential survival curve to time-to-endpoint data from placebo subjects, limited to time prior to switching to commercial use of OCA. If the re-computed sample size is less than 428, the original sample size of 428 subjects will be used. If the conditional power is less than 70% and the hazard ratio is greater than 0.85, the DMC may request additional analyses to evaluate the

feasibility of continued conduct of the study as designed. Additional details on the sample size recalculation may be found in Section 7.5.

6.   DEVIATIONS FROM PROTOCOL

6.1.   Number of Interim Analyses

Although the protocol did originally plan for two interim analyses, one at ~50% (64 events) and the other at ~75% (96 events) information fraction, the first interim analysis was postponed to occur after regulatory authority review of the interim SAP. As a result, it is anticipated that the first interim analysis will occur after the accrual of approximately 65% (83 events) of the required total clinical outcome events. The Sponsor proposed and the regulatory authority agreed that the second interim analysis be omitted. Regardless of the number of interim analyses, O'Brien-Fleming Boundaries will be used to determine the criteria for superiority. Table 2 provide the relevant estimates both for the original two-interim analyses and revised single-interim analysis scenarios.

Table 2:   Stopping Criteria Based on O'Brien-Fleming Boundaries

Type 1 Error	Power	Information Fraction	Number of Clinical Events	Incremental α Spend	Cumulative α Spend	Reject for Efficacy if Hazard Ratio
Two Interim Analyses						
		50%	64	0.003	0.003	≤0.473
0.05	0.80	75%	96	0.018	0.019	≤0.611
		End of Study (100%)	127	0.044	0.050	≤0.695
Single Interim Analysis						
		65%	83	0.011	0.011	≤0.564
0.05	0.80	End of Study (100%)	127	0.047	0.050	≤0.697

6.2.   Additional Sensitivity and Efficacy Analyses

As OCA is commercially available to many subjects enrolled in the study, informative censoring and its potential to induce bias in the primary efficacy analysis is a concern. To inform the DMC in its recommendation on the continued conduct of the study, the following additional sensitivity analyses will be conducted:

1. The hazard ratio will be estimated using Inverse Probability of Censoring Weighting (IPCW) techniques (Robins 1999). The DMC will use this analysis to assess the impact of censoring on the hazard ratio.

2. The primary efficacy analysis will be repeated using only data accrued prior to subjects discontinuing regularly scheduled study visits. The objective of this analysis is to assess whether the time subjects spend off-treatment/off study-visits is diluting the hazard ratio.

3. The primary efficacy analysis will be repeated using only data accrued prior to subjects starting commercial use of OCA. The objective of this analysis is to assess whether use of commercial use of OCA impacts the hazard ratio.

Additional efficacy analyses of correlation of time-to-event outcomes and biochemical response will be run as supportive information for primary endpoint evaluation and in support of FDA PMR 3057-3.

In addition, the Sponsor will evaluate the viability of including progression to cirrhosis in the primary efficacy endpoint in a blinded fashion (to be described in a separate document).

6.3.   Study Discontinuation

To further inform the DMC on the potential existence of informative censoring, subject discontinuation information, which is already available for DMC review, will be further detailed by baseline disease severity. The DMC will consider observed differences in censoring rates by treatment group and baseline disease severity or lack thereof when making a recommendation regarding the feasibility to conduct of the study.

6.4.   Sample Size Recalculation

Due to higher-than-expected study and treatment discontinuation rates and slower-than-expected enrollment rates, a sample size recalculation has been added in order to re-assess the assumptions for the sample size and power evaluation. The observed hazard ratio at data cut-off of the interim analysis will guide the assessment of the required number of events. Evaluation of the sample size needed to reach the required number of events will be based on the observed placebo hazard rate at data cut-off of the interim analysis, assuming a total study duration of 12 years allowing for 10 years accrual and 2 years of follow-up.

7.   APPENDIX: GENERAL ANALYSIS RULES AND DATA CONVENTIONS

The general analysis rules and data conventions described herein that are pertinent to the interim analysis will be followed.

Individual subject data obtained from electronic case report forms (eCRFs), central laboratories, external sources, and any derived data will be presented in data listings by subject. All data listings that contain an evaluation date will contain a relative study day. Pre-treatment and on-treatment study days are numbered relative to the day of the first dose of investigational product which is designated as Day 1. The preceding day is Day -1, the day before that is Day -2, etc. The last day of investigational product is designated with an "L" (eg, Day 14L). Post-treatment study days are numbered relative to the last dose and are designated as Day 1P, Day 2P, etc.

All output will be incorporated into Microsoft Word rich text format (.rtf) files, sorted and labeled according to the ICH recommendations, and formatted to the appropriate page size(s).

Tabulations will be produced for appropriate demographic, Baseline, efficacy, and safety parameters.

For categorical variables, summary tabulations of the number and percentage of subjects within each category (with a category for missing data) of the parameter will be presented. Percentage calculations will be based on non-missing data, unless otherwise specified. Percentages are rounded to 1 decimal place, unless otherwise specified.

For continuous variables, the number of subjects, mean, standard deviation (SD), standard error of the mean (SEM), median, first and third quartiles (Q1 and Q3), minimum, and maximum values will be presented. Other summaries (eg, quartiles, 5%, 95% intervals) may be used as appropriate. The precision of summary statistics, unless otherwise specified, will be as follows: mean and median to 1 more decimal place than the raw data, and SD and SEM to 2 decimal places more than the raw data. In general, the decimal places should not exceed 3 decimal places unless appropriate. Confidence intervals (CIs) will be provided and will be rounded to 1 decimal place, unless otherwise specified, in the table and listing shell.

Time-to-event data will be summarized using Kaplan-Meier methodology using 25th, 50th (median), and 75th percentiles with associated 2-sided 95% CIs, as well as percentage of censored observations.

All statistical tests comparing groups will be conducted at the 2-sided, 0.05 level of significance, unless otherwise specified (eg, the primary efficacy analysis, Section 4.2.1). Summary statistics for each treatment group will be presented, as well as 95% CIs comparing groups.

7.1. Baseline Definitions

The Baseline value for statistical analyses of quantitative parameters is defined as the mean of all available study evaluations after the subject signs the informed consent and prior to the first administration of investigational product, unless otherwise specified. If there is only one evaluation prior to the first administration of investigational product then the available data from this evaluation will be used as the Baseline value.

Baseline values defined above will not change regardless if a subject stops taking investigational product and begins taking commercially marketed OCA.

7.2. Partial Dates

If only a partial date is available and is required for calculation, the following standards will be applied:

- Diagnostic Date (eg, PBC diagnostic date)
  - For missing day only – Day will be imputed as the first day of the month (ie, 1).
  - For missing day and month – Day and month will be imputed as the first day of the year (ie, 1 January).
- Start Dates (eg, event date, AE onset date, or start date of medication)
  - For missing start day only – Day will be imputed as the first day of the month (ie, 1) with the following exception: if the partial date falls in the same month and year as the date being used in the calculation (eg, first dose date, informed

consent date), then the partial date will be imputed to equal the date being used for the calculation.

- For missing start day and month – Day and month will be imputed as the first day of the year (ie, 1 January) with the following exception: if the partial date falls in the same year as the date being used in the calculation (eg, first dose date, informed consent date), then the partial date will be imputed to equal the date being used for the calculation.

- Imputed start dates must be prior to the stop date.

- Stop Dates (eg, AE resolution date or stop date of medication)

    - For missing stop day only – Day will be imputed as the last day of the month (ie, 28, 29, 30, or 31).

    - For missing stop day and month – Day and month will be imputed as the last day of the year (ie, 31 December).

    - Imputed stop dates must be on or after the start date.

All data recorded on the case report form will be included in data listings that will accompany the clinical study report.

7.3.   Data Conventions

The precision of original measurements will be maintained in summaries, when possible.

Means, medians, SEMs, and SDs will be presented with an increased level of precision, where means and medians will be presented to one more decimal place than the raw data, and the SEMs and SDs will be presented to two more decimal places than the raw data. In general, the decimal places should not exceed three decimal places, unless appropriate.

For tables where rounding is required, rounding will be done to the nearest round-off unit. For example, when rounding to the nearest integer, values $\geq$XX.5 will be rounded up to XX+1 (eg, 97.5 will round up to 98), while values <XX.5 will be rounded down to XX (eg, 97.4 will round down to 97).

Percentages based on frequency counts will be based on available data, and denominators will generally exclude missing values, unless otherwise stated. For frequency counts of categorical variables, categories whose counts are zero will be displayed for the sake of completeness. For example, if none of the subjects discontinue due to "lost to follow-up," this reason will be included in the table with a count of 0. Percentages based on frequency counts will be presented as a whole number (no decimal places), and values less than 1% will be presented as "<1%." Values less than 100% but that round up from 99.5% to 100% will be presented as ">99%."

Quantitative laboratory tests containing less than (<) and greater than (>) symbols are test results that are below and above quantifiable limits, respectively. In order to retain these values for analysis purpose, the following imputations will be done within the analysis datasets:

For laboratory test results that are below the quantifiable limit:

Imputed laboratory results = (numeric portion of the result) x 0.9.

For laboratory test results that are above the quantifiable limit:

Imputed laboratory results = (numeric portion of the result) x 1.1.

7.4. Standard Calculations

Variables requiring calculation will be derived using the following formulas:

- Time to event – The time to an event will be calculated in days as the date of the first occurrence of the event minus the date of first investigational product administration plus 1

- Days – A duration expressed in days between one date (*date1*) and another later date (*date2*) will be calculated using the following formulas:

  duration in days = date2 – date1 + 1, where date1 ≥first dose date

  duration in days = date2 – date1, where date1 <first dose date

- Months – A duration expressed in months is calculated as the number of days divided by 365.25/12 (approximately 30.4)

- Years – A duration expressed in years between one date (*date1*) and another date (*date2*) is calculated using the following formulas:

  duration in years = (date2 – date1 + 1)/365.25, where date1 ≥first dose date

  duration in years = (date2 – date1)/365.25, where date1<first dose date

- Age – Age is calculated as the number of years from the date of birth (*DOB*) to the specified date, eg, date of informed consent (*DOIC*). If the month of DOIC <month of DOB or the month of DOIC = DOB and the day of DOIC <day of DOB, then the following formula is used:

  age (years) = year of DOIC – year of DOB – 1.

  Otherwise, the following formula is used:

  age (years) = year of DOIC – year of DOB.

- Change from Baseline – Change from Baseline will be calculated as:

  Change = post Baseline value – Baseline value

- Percentage change from Baseline – Change from Baseline will be calculated as:

  Percentage change from Baseline = ([post Baseline value – Baseline value] / Baseline value) × 100

- MELD score is derived using the following formula:

  MELD = 3.78 × ln[serum bilirubin (mg/dL)] + 11.2 × ln[INR] + 9.57 × ln[serum creatinine (mg/dL)] + 6.43 × aetiology(0: cholestatic or alcoholic, 1: otherwise).

- Child Pugh classification (Noncirrhotic/A/B/C) is defined as follows: Subjects deemed noncirrhotic at baseline by the PI are captured as "noncirrhotic" regardless of baseline Child-Pugh score and those without a cirrhosis assessment or who were

deemed cirrhotic at baseline are summarized by their baseline (Screening or Day 0) Child-Pugh category.

- Rotterdam Criteria (Mild/Moderate/Severe) is defined as follows:

    Mild - Total Bilirubin $\leq$ ULN and Albumin $\geq$ LLN

    Moderate - Total Bilirubin > ULN and Albumin $\geq$ LLN or

    Total Bilirubin $\leq$ ULN and Albumin < LLN

    Severe – Total Bilirubin > ULN and Albumin < LLN, where baseline values of total bilirubin and albumin are taken to be the average of all pre-dose measurements.

7.5. Sample Size Recalculation

With a single interim analysis at 65% information fraction, the interim analysis and final study analysis are anticipated to occur at 83 and 127 events, respectively. The pre-specified weights that will be used for the sample size re-assessment will be:

$$w_1 = \frac{D_1}{127}; \ w_2 = \frac{127 - D_1}{127}$$

D_1 is the number of actual observed events at the interim analysis.

Let

- $\delta = \ln(HR)$
- b_1 and b_2 denote the α-level stopping boundaries at the planned interim and final analysis, respectively
- r denote the proportion of subjects in the OCA treatment group
- $Z^*_{j;cum}$ denote the Z-score based on the stratified logrank statistic at look j
- D_1^* and D_2^* denote the altered cumulative events at the interim and final analyses resulting from an adaptation of the original design
- $\hat{\delta}_j$ and $\widehat{se}(\hat{\delta}_j)$ denote the estimates of treatment effect and standard error of treatment effect from fitting the Cox proportional hazard model to the cumulative data at look j

The CHW statistic for the interim ($j = 1$) and final ($j = 2$) analyses are:

$$Z^*_{1,CHW} = \frac{\sqrt{w_1}Z^{*(1)}}{\sqrt{w_1}} \text{ and } Z^*_{2,CHW} = \frac{\sqrt{w_1}Z^{*(1)} + \sqrt{w_2}Z^{*(2)}}{\sqrt{w_1 + w_2}}, \text{respectively,}$$

where $Z^{*(j)}$ is the incremental statistic at look j:

$$Z^{*(j)} = \frac{\sqrt{I_j^*}Z_{j,cum}^* - \sqrt{I_{j-1}^*}Z_{j-1,cum}^*}{\sqrt{I_j^* - I_{j-1}^*}}$$

and $I_j^* \approx \dfrac{1}{\left(\widehat{se}(\hat{\delta}_j)\right)^2}$

If the CHW statistic at the interim analysis is $Z_{1,CHW}^* = z_1$ and D_2^* cumulative events are required at the final analysis, the conditional power at the interim analysis is:

$$CP_\delta(z_1) = 1 - \Phi\left\{b_2\sqrt{1 + \frac{D_1}{D_2 - D_1}} - z_1\sqrt{\frac{D_1}{D_2 - D_1}} - \delta\sqrt{r(1-r)}\sqrt{D_2^* - D_1}\right\}$$

Where the log hazard ratio δ will be estimated using $\hat{\delta}_1$, the estimate of treatment effect from fitting the Cox proportional hazard model to the cumulative data at the interim analysis.

8.   REFERENCES

Cui L, Hung HM, Wang SJ. Modification of Sample Size in Group Sequential Clinical Trials. Biometrics. 1999. Sep;55(3):853-857

Lan KKG and DeMets DL. Discrete Sequential Boundaries for Clinical Trials. Biometrika. 1983;70:659-663.

Reboussin DM, DeMets, DL, Kim KM, et al. Computations for Group Sequential Boundaries Using the Lan-DeMets Spending Function Method. Controlled Clin Trials. 2000;21(3):190-207.

Robins JM. Association, Causation, and Marginal Structural Models. Synthese. 1999 Jan 1:151-179.

9. REVISION HISTORY

9.1. Rationale of the Revision

IND 063307 Advice-Information Request was received from the agency after the submission of the SAP v1.0, dated 14 July 2020. The SAP was amended in response to the agency's comments.

9.2. Changes from Previous Version

The following changes have been implemented in this SAP amendment:

Section	Original Text (Version 1.0, 14 Jul 2020)	Revised Text (Amendment 1.0, 20 Aug 2020) in Red
1.1.1 1.2.2 4.2.1	– Uncontrolled ascites (diuretic resistant ascites requiring therapeutic paracentesis at a frequency of at least twice in a month)	• Format changes for this bullet point as this is one component of the composite event endpoint and should not be under the hospitalization.
4.1	Group sequential boundaries for efficacy at the planned timepoints are provided in Table 1 for the Type I error rate comprising the sum of prespecified alpha for clinical outcomes and carried forward alpha from the interim analysis. Exact boundaries and alpha levels will be calculated at the time of analysis based on the actual observed information fraction and observed discontinuation rates. The criteria for the DMC to recommend that the study stop for efficacy will be met at the interim analysis when the estimated hazard ratio is less than δ^*, where δ^* is the hazard ratio efficacy boundary calculated based on observed data at the interim analysis.	Stopping criteria for efficacy at the planned timepoints based on group sequential boundaries are provided in Table 1 for the Type I error rate comprising the sum of prespecified alpha for clinical outcomes and carried forward alpha from the interim analysis. The exact alpha level and corresponding hazard ratio criterion at the time of interim analysis will be calculated based on the actual observed number of clinical events. The criteria for the DMC to recommend that the study stop for efficacy will be met at the interim analysis when the estimated hazard ratio is less than δ^*, where δ^* is the hazard ratio efficacy criterion calculated based on observed data at the interim analysis.

Section	Original Text (Version 1.0, 14 Jul 2020)	Revised Text (Amendment 1.0, 20 Aug 2020) in Red
4.1	Table 1: O'Brien-Fleming Boundaries Note: Estimates for information fraction and number of clinical events at interim analysis are provided in Table 1 for illustration, and the hazard ratio efficacy boundary δ^* will be calculated based on observed data.	Table 1: Stopping Criteria Based on O'Brien-Fleming Boundaries Note: Estimates for information fraction and number of clinical events at interim analysis are provided in Table 1 for illustration, and the hazard ratio efficacy criteria δ^* will be calculated based on observed data.
4.2.2	1. The hazard ratio will be estimated using Inverse Probability of Censoring Weighting (IPCW) techniques (Robins 1999). The DMC will use this analysis to assess the impact of censoring on the hazard ratio. Baseline predictors of both censoring and clinical outcome will be identified in advance. Augmented IPCW estimators described by Robins may be implemented to explore the possible impact of unmeasured confounding.	1. The hazard ratio will be estimated using Inverse Probability of Censoring Weighting (IPCW) techniques (Robins 1999). The DMC will use this analysis to assess the impact of censoring on the hazard ratio. Baseline predictors of both censoring and clinical outcome will be identified in advance. Augmented IPCW estimators described by Robins may be implemented to explore the possible impact of unmeasured confounding. Baseline predictors of both censoring and clinical outcome for the IPCW analysis are: • Treatment Group: OCA/Placebo • Age (year) at Screening Visit • Sex: Male/Female • UDCA use at Screening Visit: Yes/No • Baseline Total Bilirubin (µmol/L) • Baseline Alanine Transaminase (U/L) • Baseline Aspartate Transaminase (U/L)

Section	Original Text (Version 1.0, 14 Jul 2020)	Revised Text (Amendment 1.0, 20 Aug 2020) in Red
		• Baseline Alkaline Phosphatase (U/L) • Baseline Albumin (g/L) • Baseline Platelets count The post-baseline time-varying covariates are: ALP, ALT, AST, Total Bilirubin, and AE of pruritus with severity of moderate and severe. To derive the IPCW weights, the patient's follow-up time up until the time of censoring or event occurring will be partitioned into about 10 intervals. The probability of remaining uncensored at the end of each time interval adjusted for baseline variables, and post-baseline time-varying covariates will be estimated using a pooled logistic regression model. To avoid possible extreme values when taking the inverse of the estimated probabilities from the pooled logistic model with both baseline and time-varying post-baseline covariates, the inverse of these probabilities will be stabilized by multiplying the probability of remaining uncensored conditional only on baseline covariates. Once the IPCW weights are derived, the HR of treatment effect and corresponding 95% CI will be estimated using a weighted Cox regression model with Sandwich estimator to obtain the robust estimate of the variance-covariance matrix of the parameter estimates.

Section	Original Text (Version 1.0, 14 Jul 2020)	Revised Text (Amendment 1.0, 20 Aug 2020) in Red
5.2	If the criteria for the DMC to recommend that the study stop for efficacy, ie, the estimated hazard ratio is less than δ^*, where δ^* is the hazard ratio efficacy boundary at the interim analysis, are not met at the interim analyses, the conditional power for rejecting the null hypothesis in favor of the alternative (ie, the probability of success at end of trial given the interim results) will be computed.	If the criteria for the DMC to recommend that the study stop for efficacy, ie, the estimated hazard ratio is less than δ^*, where δ^* is the hazard ratio efficacy criterion at the interim analysis, are not met at the interim analyses, the conditional power for rejecting the null hypothesis in favor of the alternative (ie, the probability of success at end of trial given the interim results) will be computed.
6.1	Table 2: O'Brien-Fleming Boundaries	Table 2: Stopping Criteria Based on O'Brien-Fleming Boundaries
7.5	With a single interim analysis at 65% information fraction, the interim analysis and final study analysis are anticipated to occur at $D_1 = 83$ and $D_2 = 127$ events, respectively.	With a single interim analysis at 65% information fraction, the interim analysis and final study analysis are anticipated to occur at 83 and 127 events, respectively. D_1 is the number of actual observed events at the interim analysis.
9	N/A	Revision History section added