Statistical Analysis Plan
For the Long-Term Safety Extension of
Protocol 747-207

A Phase 2, Randomized, Double-Blind, Placebo-Controlled, Dose-Finding, Clinical Trial Evaluating the Efficacy and Safety of Obeticholic Acid in Subjects with Primary Sclerosing Cholangitis

OBETICHOLIC ACID (OCA)

Protocol Version and Date:	Version 5: 18 March 2016
Phase:	Phase 2
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
Analysis Plan Date:	18 May 2018
Analysis Plan Version:	Final Version 1

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation or Specialist Term	Explanation
AE	adverse event
AESI	adverse event of special interest
ALP	alkaline phosphatase
ALT	alanine transaminase
ANCOVA	analysis of covariance
Apo	apolipoprotein
ApoA1	apolipoprotein A-1
ApoB	apolipoprotein B
ApoE	apolipoprotein E
AST	aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
AUC_{0-6}	area under the concentration-time curve from hour 0 to last sampling time (hour 6)
AUC_t	area under the concentration time curve
BLQ	below the limit of quantitation
BMI	body mass index
BP	blood pressure
BUN	blood urea nitrogen
C4	7α-hydroxy-4-cholesten-3-one
CA	cholic acid
CDAI	Crohn's disease activity index
CDCA	chenodeoxycholic acid
CI	confidence interval
CK-18	cytokeratin 18
cm	centimeter(s)
C_{max}	maximum plasma concentration
CSR	clinical study report
C_{ss}	steady state concentration
CV	coefficient of variation
DB	double-blind

Abbreviation or Specialist Term	Explanation
DCA	deoxycholic acid
DOB	date of birth
DOIC	date of informed consent
eCRFs	electronic case report forms
ECG	Electrocardiogram
eCRF	electronic case report form
ELF	enhanced liver fibrosis
eq	equivalents
ERCP	endoscopic retrograde cholangiopancreatography
FGF-19	fibroblast growth factor-19
FXR	farnesoid X receptor
g	gram(s)
GGT	gamma-glutamyl transferase
GI	Gastrointestinal
H	High
HA	hyaluronic acid
IBD	Inflammatory bowel disease
ICF	informed consent form
ICH	International Conference on Harmonization
IgA	Immunoglobulin A
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IL	interleukin
INR	international normalized ratio
in	inch(es)
ITT	Intent-to-treat

Abbreviation or Specialist Term	Explanation
kg	kilogram(s)
KM	Kaplan-Meier
L	low
lb	pound(s)
LCA	lithocholic acid
LLN	lower limit(s) of normal
LLQ	lower limit of quantitation
ln	natural logarithm
LOCF	last observation carried forward
LS	least-square
LTSE	long-term safety extension
MedDRA	Medical Dictionary for Regulatory Activities
mg	milligram(s)
mL	milliliter
MRCP	magnetic resonance cholangiopancreatography
MW	molecular weight
ng	nanogram
NMR	nuclear magnetic resonance
OCA	obeticholic acid
ODS	output delivery system
P3NP	procollagen-3 N-terminal peptide
PD	pharmacodynamic
PK	pharmacokinetic(s)
PSC	primary sclerosing cholangitis
SAE	serious adverse event
SAP	statistical analysis plan
SD	standard deviation
SE	standard error
SEM	standard error of mean
SI	international system of units
tauro-CA	taurine 6α-ethyl cholic acid
tauro-CDCA	taurine 6α-ethyl chenodeoxycholic acid

Abbreviation or Specialist Term	Explanation
tauro-DCA	taurine 6α-ethyl deoxycholic acid
tauro-LCA	taurine 6α-ethyl lithocholic acid
tauro-OCA	taurine 6α-ethyl chenodeoxycholic acid
tauro-UDCA	taurine 6α-ethyl ursodeoxycholic acid
TE	transient elastography
TEAE	treatment-emergent adverse event
TGF-β	transforming growth factor-beta
TIMP-1	tissue inhibitor of metalloproteinase 1
t_{max}	time to reach C_{max}
TNF-α	tumor necrosis factor alpha
UC	ulcerative colitis
UDCA	ursodeoxycholic acid
ULN	upper limit(s) of normal
VAS	visual analogue scale
WADD	Weighted average daily dose
WHODDE	World Health Organization Drug Dictionary Enhanced

1. INTRODUCTION

Subjects who complete the 6-month, double-blind phase of Study 747-207 (A Phase 2, Randomized, Double-Blind, Placebo-Controlled, Dose-Finding, Clinical Trial Evaluating the Efficacy and Safety of Obeticholic Acid in Subjects with Primary Sclerosing Cholangitis), conducted by Intercept Pharmaceuticals, Inc. (the Sponsor), have the option to participate in the open-label long-term safety extension (LTSE) phase. Data from the LTSE provide a thorough evaluation of the long-term safety and efficacy of obeticholic acid (OCA) in subjects with primary sclerosing cholangitis (PSC), including durability of response.

This statistical analysis plan (SAP) outlines the statistical methods to be implemented during the analyses of the LTSE phase of Study 747-207. The purpose of this plan is to provide specific guidelines from which the analyses will proceed. Any deviations from this plan will be documented in the clinical study report (CSR). The scope of this plan includes the detailed specifications of the statistical analyses for the LTSE phase only. The statistical analysis plan (SAP) was submitted to file during the conduct of the open-label safety extension (LTSE) phase, before database lock. The analysis for the double-blind phase of the study is described in a separate SAP dated 05 April 2017.

2. STUDY OBJECTIVES

The primary objective of the LTSE phase of Study 747-207 is to evaluate the safety of OCA.

The secondary objectives are to evaluate the effects of OCA in subjects with PSC on:

- Serum alkaline phosphatase (ALP)
- Hepatic biochemistry and indices of function
- Markers of:
  - Hepatic fibrosis and gastrointestinal (GI) inflammation and disease
  - Farnesoid X receptor (FXR) activity
  - Inflammatory bowel disease (IBD)
- Bile acids
- Disease-specific symptoms

3. STUDY DESIGN AND PLAN

This is a Phase 2, randomized, double-blind, placebo-controlled, dose-finding evaluation of the efficacy and safety of OCA in subjects with PSC, followed by an LTSE phase. The trial design is shown in Figure 1. A total of 77 subjects were randomly allocated to treatment with placebo (n = 25), 1.5 mg OCA titrated to 3 mg OCA (n = 26), or 5 mg OCA titrated to 10 mg OCA (n = 26) and 76 subjects received at least 1 dose of investigational product. One subject

was randomized to the placebo group but received 5 mg OCA titrated to 10 mg OCA instead of placebo, and 1 subject allocated to the 1.5 mg-3 mg OCA group withdrew consent prior to dosing. Therefore, 24 subjects received placebo, 25 subjects received 1.5 mg OCA titrated to 3 mg OCA, and 27 subjects received 5 mg OCA titrated to 10 mg OCA. Subjects administered investigational product orally, once daily for 2 consecutive 12-week periods. Randomization was stratified by the presence or absence of concomitant UDCA use and total bilirubin level (≤1.5×ULN or >1.5×ULN but <2.5×ULN). No more than 50% of subjects randomized will be administering UDCA at the time of randomization.

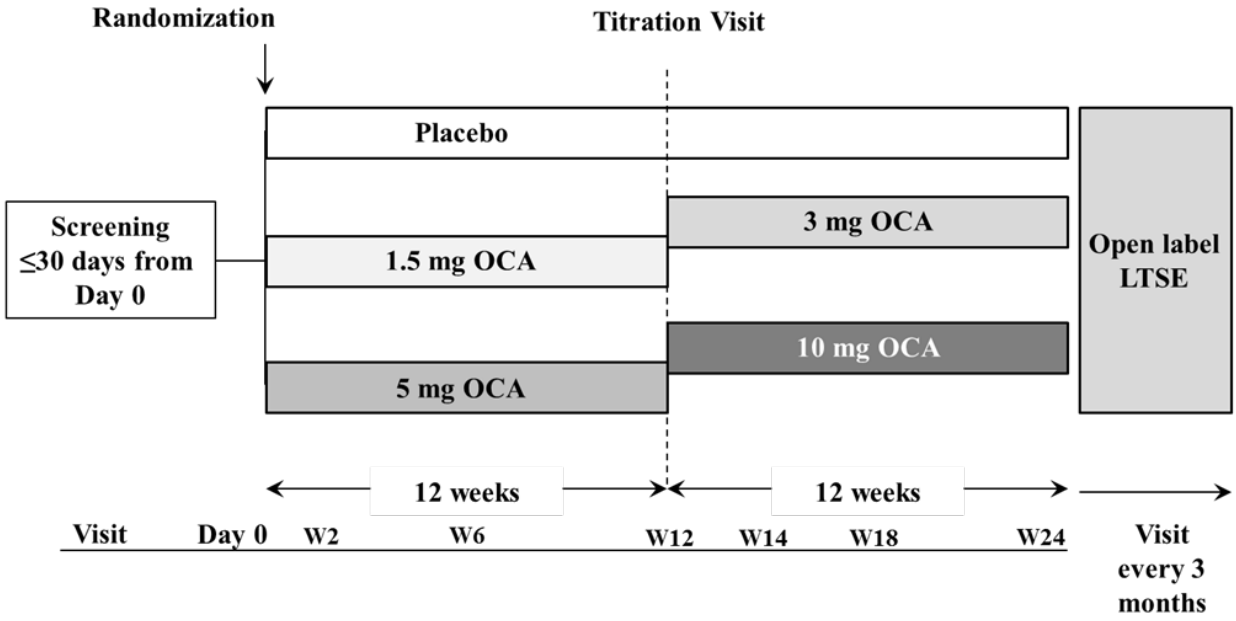
For the first 12 weeks after randomization, the subject's dose was 1.5 mg OCA, 5 mg OCA, or placebo. After 12 weeks, if the initial dose was tolerated, the subject's dose was increased – 1.5 mg OCA titrating to 3 mg OCA and 5 mg OCA titrating to 10 mg OCA – and double-blind (DB) treatment continued for a further 12 weeks at that dose.

Subjects had a screening period of up to 30 days prior to Randomization/Day 0. Subjects attended on site clinic visits at DB Weeks 2, 6, 12, 14, 18, and 24. The final visit during the DB phase occurred at Week 24, after which subjects were asked to reconfirm their consent for participation in the LTSE phase (a further 24 months).

Upon a subject's completion of the DB phase, the trial blind was broken in order to assign the starting OCA dose for the LTSE phase. It was intended that subjects would commence treatment at 5 mg OCA, except (1) the subjects who completed treatment in the DB phase with 10 mg OCA who would continue at 10 mg OCA unless safety and tolerability warranted a dose reduction to 5 mg; (2) the subjects who did not up-titrate their 1.5 mg OCA dose at Week 12 in the DB phase and remained on 1.5 mg OCA; or (3) other doses approved by the Intercept Medical Monitor. The titration schedule and options for the LTSE phase are detailed in the protocol.

The overall study duration is up to 32 months, including up to a 30-day Screening period, 6-month DB period, followed by a 24-month open-label LTSE period, and a 1-month follow-up period. Subjects were to be contacted by study site staff 2 weeks after starting the LTSE phase. Clinic visits during LTSE were to occur every 3 months (± 2 weeks), with assessments conducted as outlined in the Schedule of Trial Procedures (Protocol 747-207, Section 6.1.1, Table 2).

Figure 1: Study Design Schematic



4. DETERMINATION OF SAMPLE SIZE

No formal sample size calculations were performed for the LTSE phase of the study as this is a continuation of the double-blind phase; all subjects who complete treatment during the double-blind phase are eligible to continue. Sample size considerations for the double-blind phase can be found in the double-blind phase SAP (Version 2: 05 April 2017).

5. STUDY ENDPOINTS

5.1. Primary Endpoints

5.1.1. Primary Efficacy Endpoints

No primary efficacy endpoints are defined for LTSE.

5.2. Secondary Endpoints

The secondary endpoints for LTSE are ALP, indices of hepatic function and GI inflammation, hepatic biochemistry, hepatic inflammation and fibrosis, PD bile acids, and disease specific symptoms. The secondary efficacy endpoints are described below.

5.2.1. ALP

Observed values, change from DB Baseline, and percentage change in ALP from DB Baseline will be evaluated at each post-Baseline visit.

5.2.2. Hepatic Biochemistry and Indices of Hepatic Function

The following laboratory parameters will be summarized for hepatic biochemistry and indices of hepatic function: albumin, alanine transaminase (ALT), aspartate aminotransferase (AST), total bilirubin, direct bilirubin, gamma-glutamyl transferase (GGT), international normalized ratio (INR).

- Observed values, change from DB Baseline, and percentage change from DB Baseline will be evaluated at each post-Baseline visit

5.2.3. FXR Activity

FXR activity will be assessed by measuring fibroblast growth factor-19 (FGF-19) and 7-alpha-hydroxy-4-cholesten-3-one (C4).

- Observed values, change from DB Baseline, and percentage change from DB Baseline will be evaluated at each post-Baseline visit

5.2.4. Disease Specific Symptoms

The disease specific endpoints include the following: Pruritus Visual Analogue Scale (VAS), 5-D Itch questionnaire, partial Mayo score assessment for subjects with ulcerative colitis (UC), and Crohn's disease activity index (CDAI).

- Observed values and change from DB Baseline will be evaluated at each post-Baseline visit

6. GENERAL ANALYSIS CONSIDERATIONS

6.1. Data Reporting

The statistical analyses will be reported using summary tables, figures, and data listings. No statistical comparisons between treatment groups will be performed. Within-treatment group comparisons will be made where indicated.

Laboratory units will be summarized and presented both in conventional units (CV) and in the international system of units (SI).

Individual subject data obtained from the electronic case report forms (eCRFs) and external laboratory data will be presented in data listings by subject. Data from all assessments, whether scheduled or unscheduled, will be listed by subject and visit. Unscheduled visits and visits occurring more than one day outside protocol defined window will not be included in the table summaries, with exception of transient elastography measurements, which could be assessed up to 5 days outside protocol defined visit window.

The analyses described in this plan are considered a priori, in that they have been defined prior to database lock. Post hoc analyses will be labeled as such on the outputs and identified in the CSR.

All analyses and tabulations will be performed using SAS® Version 9.2 or higher. The following processes will be employed to validate statistical outputs: derived datasets (both tabulation

datasets and analysis datasets), summary tables, and data listings will be verified through independent programming; graphical displays will be compared against supporting summary tables; and all outputs will undergo a senior-level statistical review. The process includes confirmation that statistically valid methods have been implemented and that all data manipulations and calculations are appropriate and accurate. Checks will be made to ensure accuracy, adherence to this SAP, consistency within tables, and agreement between tables and their corresponding data listings. Upon completion of validation/verification and quality review procedures, all documentation will be collected and filed in the study master file by the project statistician or designee.

6.2. Data Analysis and Summaries

6.2.1. Arithmetic Summaries

Continuous variables will be summarized by means, standard deviations (SDs), standard errors of the mean (SEMs), medians, first quartile and third quartiles, minimums, and maximums.

6.2.2. Categorical Methods

Categorical variables will be summarized by counts and percentage of subjects in corresponding categories. Percentages will be based on the number of non-missing assessments unless otherwise specified.

6.2.3. Multiple Comparisons/Multiplicity

As analyses performed within the LTSE are considered descriptive and exploratory, no adjustments for multiplicity will be made for LTSE analyses.

6.2.4. Subgroup Analyses

Subgroup analyses, similar to those performed for the double-blind phase, may be performed for the LTSE phase as deemed appropriate.

6.3. Data Handling

6.3.1. Baseline Values

Baseline values are defined as follows:

- **Efficacy, Clinical laboratory, and PD endpoints**: the mean of all available evaluations prior to first dose of double-blind investigational product.
- **ECG assessment**: the last available non-missing assessment prior to first dose of double-blind investigational product.

For any other quantitative parameters, Baseline is defined as the mean of all available evaluations prior to administration of investigational product on Day 1 of DB phase.

6.3.2. Missing Data

Analysis will be done using observed data only; missing values will not be imputed.

6.3.3. Partial Dates

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

- Diagnosis date (eg, PSC diagnosis date)
  - For missing day only: Day will be imputed as the first day of the month (ie, 01).
  - For missing day and month: Day and month will be imputed as the first day of the year (ie, 01 January).
- Start dates (eg, AE onset date or start date of concomitant medication)
  - For missing start day only: Day will be imputed as the first day of the month (ie, 01) with the following exception: if the partial date was recorded on the double-blind CRF and 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; if the partial date was recorded only on the LTSE CRF and falls in the same month and year as the date of the first LTSE dose, then the date is imputed to the date of the first LTSE dose.
  - For missing start day and month: Day and month will be imputed as the first day of the year (ie, 01 January) with the following exception: if the partial date was recorded on the double-blind CRF and 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; if the partial date was recorded only on the LTSE CRF and falls in the same year as the date of the first LTSE dose, then the date is imputed to the date of the first LTSE dose.
  - 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 dates should not extend beyond the date of completion or discontinuation of the LTSE phase or the data cutoff date.
  - Imputed stop dates must be on or after the start date.

6.3.4. Data Conventions

Unless otherwise specified, in summary tables of continuous variables, the minimum and maximum values will be displayed to the same number of decimal places as the raw data, the mean, median, and quartiles will be presented to one extra decimal place compared to the raw data, and the SD and SE of the mean will be displayed to two extra decimal places compared to

the raw data. In general, the decimal places should not exceed 3 decimal places, unless appropriate.

Rounding will only occur after all calculations have been incorporated. 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 (eg, 97.5 will round up to 98), whereas values <XX.5 will be rounded down to XX (eg, 97.4 will round down to 97).

Percentages based on frequency counts of categorical variables will be based on available data, and denominators will generally exclude missing values unless some form of imputation is defined or a 'missing' category is presented. For frequency counts of categorical variables, categories with zero counts 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 for data without pre-determined categories, such as adverse events, will use the total number of subjects in the corresponding treatment group as denominator. Percentages based on frequency counts will be presented as a whole number (no decimal places). Values greater than 0 and less than 1% will be presented as "<1%", and values equal to or greater than 99.5% but less than 100% will be presented as ">99%".

6.3.5. Standard Calculations

Variables requiring calculation will be derived using the following formulas:

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

  duration (days) = date2 – date1 + 1

- **Months**: A duration expressed in months is calculated as the number of days divided by 365.25 / 12.

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

  duration (years) = (date2 – date1+ 1) / 365.25

- **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.

- **Height**: Height measured in inches (in) are converted to centimeters (cm) using the following formula:

  height (cm) = height (in) × 2.54

- **Weight**: Weight measured in pounds (lb) are converted to kilograms (kg) using the following formula:

weight (kg) = weight (lb) / 2.2046

- **Temperature**: Temperature measured in degrees Fahrenheit are converted to degrees centigrade using the following formula:

  temp (degrees centigrade) = 5 / 9 × (temp [degrees Fahrenheit] – 32)

- **Body Mass Index (BMI)**: BMI is calculated using height (cm) and weight (kg) using the following formula:

  BMI (kg/m^2) = weight (kg) / ([height (cm)/100]^2)

- **Change from Baseline**: Change from Baseline is calculated as:

  change from Baseline = post-Baseline value – DB Baseline value

- **Percentage change from Baseline**: The denominator for calculating percentage change from baseline will be the absolute value of baseline, so that the percentage change will have the same sign as the change from baseline value when the baseline value is negative, such as for CDAI. Percentage change from Baseline is calculated as:

  percentage change from Baseline = 100 × ([post-Baseline value – DB Baseline value] / absolute value of DB Baseline value)

6.3.6. Pruritus 5-D Itch Score Calculations

The 5-D questionnaire (Elman 2010) is a specific tool used to quantify the magnitude of pruritus in PBC (and other diseases). It consists of 5 domains: duration, degree, direction, disability, distribution, as well as the total score. The following rules will be used to calculate the domain scores:

- Single item domain scores (duration, degree, and direction) are equal to the value indicated below the response choice (range from 1-5).
- The disability domain includes 4 items that assess the impact of itching on daily activities: sleep, leisure/social activities, housework/errands, and work/school. The score for the disability domain is achieved by taking the highest score on any of the 4 items. The disability domain will only be calculated if at least 3 daily activities are documented, otherwise this domain score will be set to missing.
- For the distribution domain, only the section "Mark whether itching has been present in the following parts of your body over the last 2 weeks" will be used. The distribution domain includes 16 potential locations of itch, including 15 body part items and one point of contact with clothing. Points of contact with clothing will only count as one point. Missing ticks are interpreted to be absent. The number of affected body parts ('present') is tallied (potential sum 0 to 16) and the sum is sorted into 5 scoring bins: sum of 0 to 2 = score of 1, sum of 3 to 5 = score of 2, sum of 6 to 10 = score of 3, sum of 11 to 13 = score of 4, sum of 14 to 16 = score of 5.
- The total 5D score is obtained by summing up the domain scores and ranges between 5 (no pruritus) and 25 (most severe pruritus). The total score will not be calculated if any of the domain scores is missing.

- If multiple answers are given for any item then the most severe one will be used.

7. ANALYSIS POPULATIONS

The following analysis population will be used:

7.1. Long-Term Safety Extension (LTSE) Population

The LTSE population will include all subjects who receive any amount of investigational product during the long-term safety extension (LTSE) phase. Treatment assignment will be based on the treatment actually received during the DB phase. The LTSE population will be used for all safety and efficacy analysis.

8. STUDY POPULATION

8.1. Subject Disposition

Subject disposition information will be summarized and listed for all LTSE subjects. Summaries will include the following: the number of subjects completing the LTSE phase (Month 24) and each scheduled visit (Month 3, 6, 9, 12, 15, 18, and 21), and the primary reason for discontinuation from the LTSE phase. A subject will be considered as completing the LTSE phase (Month 24) and each scheduled visit (Month 3, 6, 9, 12, 15, 18, and 21) if the subject has a non-missing assessment of ALP at the corresponding visit. Percentages will be based on the LTSE Population.

8.2. Protocol Deviations

Protocol deviations for missed visits, missed assessments, out of window visits or assessments, and violations of inclusion/exclusion criteria will be determined based on available data. All other protocol deviations will be collected by the clinical research associates. Major protocol deviations identified during the study that could potentially affect the conclusions of the study or result in a subject's removal from an analysis population will be classified as such prior to database lock. Major protocol deviations will be summarized by deviation category and treatment group. All protocol deviations and their classifications will be presented in a listing. Subjects with protocol deviations that result in a subject's removal from analysis will be flagged in a listing.

8.3. Demographic and Double-Blind Baseline Characteristics

Demographic variables will include the following:

- Age at informed consent
- Age at informed consent categorized as <65 years and ≥65 years
- Sex
- Race/Ethnicity

Other Baseline characteristics will include the following:

- Weight (kg)
- Height (cm)
- BMI (kg/m^2)
- BMI categorized as <30 kg/m^2 and ≥ 30 kg/m^2
- Hepatic ultrasound to assess bile duct patency performed (Yes/No)
- UDCA Use (Yes/No)
- Total daily dose of UDCA (mg/kg) as determined using the last UDCA dosage reported prior to IP dosing in the double-blind phase divided by the weight in kg at baseline.
- Key liver function test results as a continuous variable and categorized as follows by the ULN and the LLN unless otherwise specified:
  - ALP: ≤ 3xULN, >3xULN; and, ≤ median baseline ALP, > median baseline ALP
  - Total Bilirubin: ≤ ULN, > ULN, > ULN - ≤1.5xULN, >1.5xULN - ≤2xULN, >2xULN - ≤2.5xULN

International normalized ratio (INR): ≤1.3, >1.3 Arithmetic summary statistics as described in Section 6.2.1 will be presented for age, weight, height, BMI, key liver function test results, and total daily dose of UDCA. Frequency counts and percentages will be presented for age groups, sex, race/ethnicity, BMI category, key liver function categories, hepatic ultrasound, and use of UDCA.

Race and ethnicity will be summarized as follows:

- Race
  - American Indian or Alaska Native
  - Asian
  - Black or African American
  - Native Hawaiian or Other Pacific Islander
  - White
- Ethnicity
  - Hispanic or Latino
  - Not Hispanic or Latino
  - Not reported
  - Unknown

Demographic and Baseline characteristics will be summarized for the LTSE population. All demographic and Baseline characteristics will be presented in data listings. The subgroup tables will be generated for IBD regardless of the total number of patients in each group.

8.4. PSC Disease History

Baseline PSC disease characteristics will be summarized using data collected from the PSC Disease History eCRF. Assessments include the following:

- Age at first PSC diagnosis
- Age at first PSC diagnosis categorized as <40 years and ≥40 years
- Symptomatic at diagnosis (Yes/No)
- Age at first occurrence of PSC symptoms
- Age at first occurrence of PSC symptoms categorized as <40 years and ≥40 years
- Duration of PSC in years at time of informed consent
- Duration of PSC categorized as ≤ median years and > median years
- PSC diagnosis confirmed via cholangiagraphy (Yes/No)
- Modality of PSC diagnosis (endoscopic retrograde cholangiopancreatography [ERCP], magnetic resonance cholangiopancreatography [MRCP], Other)
- Affected bile ducts (intra-hepatic, extra-hepatic, both, unknown)
- Symptoms of PSC (Never, Previous, Current)
  - Right upper quadrant abdominal pain
  - Fatigue
  - Weight loss
  - Jaundice
  - Pruritus
  - Other
- Severity of most recent pruritus event (Mild, Moderate, Severe)
- UDCA Use (Never/Previous/Current)
- Highest previous daily dose of UDCA (mg/kg)
- Inflammatory Bowel Disease (IBD) (No IBD, Ulcerative Colitis, Crohn's Disease, Other)

Arithmetic summary statistics as described in Section 6.2.1 will be presented for the age at first PSC diagnosis, age at first occurrence of PSC symptoms, duration of PSC, highest previous daily dose of UDCA (mg/kg). All other categorical PSC disease characteristics will be summarized using frequency counts and percentages. PSC disease history will be summarized for the LTSE population. The subgroup tables will be generated for IBD regardless of the total number of patients in each group. PSC disease history will be presented in a data listing.

8.5. General Medical History

Verbatim medical history terms on eCRFs will be mapped to preferred terms and system organ classes using the Medical Dictionary for Regulatory Activities (MedDRA) version 17.1. Frequency counts and percentages of the number of subjects reporting an abnormal Baseline medical history will be summarized by MedDRA system organ class and preferred term using the LTSE Population.

Summaries that are displayed by system organ class and preferred term will be ordered by descending order of incidence of system organ class and preferred terms within each system organ class.

Medical history data will be presented in a data listing.

8.6. Prior and New Concomitant Medications

Verbatim terms on eCRFs will be mapped to ATC class and preferred term using the World Health Organization Drug Dictionary Enhanced (WHO-DDE June 2014).

Pretreatment medications are those medications with start and stop dates prior to the first dose of investigational product in the DB phase. Prior concomitant medications are those medications that started prior to, and continued after, the first dose of investigational product in the DB phase. New concomitant medications are those medications that were started after the first dose of investigational product in the DB phase. If it cannot be determined whether the medication was a new concomitant medication due to a partial start or stop date or if the medication is taken on the same date as the first dose in the DB phase, then it will be counted as a new concomitant medication.

Pretreatment, prior, and new medications will be presented in listings. New concomitant medications will be summarized by World Health Organization ATC class and preferred term using the LTSE Population. The summary will present the number and percentage of subjects using each medication. Subjects may have more than one medication per ATC class and preferred name. A subject is counted once if one or more medication is reported at the ATC class or preferred term level. Each summary will be ordered by descending order of incidence, for all LTSE patients, of ATC class and preferred term within each ATC class.

9. EFFICACY ANALYSES

The primary and secondary efficacy analyses will be based on the LTSE population.

9.1. Primary Efficacy Analyses

There are no primary efficacy endpoints defined for LTSE.

9.2. Secondary Efficacy Analyses

All secondary efficacy endpoints will be analyzed for LTSE Population.

9.2.1. ALP

Absolute change from Baseline and percentage change from Baseline in ALP will be summarized by DB treatment group at DB baseline and at each scheduled post-Baseline visit using arithmetic summary statistics as described in Section 6.2.1. The change from Baseline results within each treatment group will be compared at two visits using a paired t-test.

9.2.2. Hepatic Biochemistry and Indices of Function

Secondary efficacy analyses of hepatic biochemistry and function parameters as described in Section 5.2.2 will be summarized by treatment group using descriptive statistics at DB Baseline and at each scheduled post-Baseline visit. The change from DB Baseline will also be summarized. The change from Baseline results within each treatment group will be compared at two visits using a paired t-test.

.

9.2.3. FXR Activity

FXR activity will be assessed by measuring FGF-19 and C4. These parameters will be summarized by DB treatment group at DB baseline and at each scheduled post-Baseline visit using arithmetic summary statistics as described in Section 6.2.1.

9.2.4. Disease-specific symptoms

The Pruritus VAS, Pruritus 5-D Itch questionnaire, Partial Mayo scoring system for assessment of UC activity (partial Mayo score), and Crohn's disease activity index (CDAI) will be summarized using descriptive statistics as described in Section 6.2.1. This analysis will be repeated for change from DB Baseline and percent change from DB Baseline in these scores.

The proportion of subjects in remission from Crohn's disease (CDAI < 150) will be summarized by treatment group and visit using statistics as described in Section 6.2.2. The analysis will be repeated for the proportions of subjects in remission from UC (partial Mayo score of ≤2 with no individual sub-score exceeding 1 point), and the proportion with mild disease (a partial Mayo score ≤3 with no individual sub-score exceeding 1 point).

10. SAFETY ANALYSES

All safety analyses will be based on the LTSE Population. The evaluation of the effects of OCA on safety parameters in the PSC population is a primary objective of this study.

10.1. Extent of Exposure

10.1.1. Investigational Product (OCA or Placebo)

During the DB period, subjects will titrate their dose of OCA or placebo at the Week 12 visit according to the criteria specified in the protocol. Upon entry into the LTSE, it is intended that subjects will commence treatment at 5 mg OCA unless the subject is already dosing >5 mg or did not increase dose to >1.5 mg OCA. During the LTSE, doses up to a maximum daily dose of 10 mg are permitted; dosing modifications based on tolerability and efficacy are specified in the

protocol. Frequency counts and percentages of the number of subjects who do not up-titrate will be summarized. The duration of investigational product exposure will be calculated as follows:

- Exposure to investigational product = {[(Date of last investigational product dose–Date of 1st investigational product dose) + 1] - Total duration of temporary investigational product 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 investigational product - Date of temporary discontinuation of investigational product.

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

Total investigational product (mg) exposed to subject will be calculated by adding the doses taken by a subject during the study and will be summarized using descriptive statistics.

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

- (# of days consumed during study)/ (# of days planned on study drug, excluding drug holidays) * 100

Weighted average daily dose (WADD) of OCA will be calculated as the sum of (dose x number of days at specified dose) / total number of days on OCA. For example, a subject on 5 mg for 6 months and 10 mg for 6 months without any days off OCA would be calculated as [(5 mg x 180 days) + (10 mg x 180 days)] / 360 days = 7.5 mg. The weighted average daily dose will also be categorized as ≤5 mg OCA and >5-10 mg OCA.

Duration of investigational product exposure, number of accidentally missed doses, exposure to investigational product, duration of temporary discontinuation of investigational product, investigational product compliance, and weighted average daily dose will be summarized by treatment group using descriptive statistics.

Additional summaries will present the number and percentage of subjects with any drug interruption, subjects with alternate dosing, subjects with reduced dose, subjects who did not titrate dose at Week 12, number of subjects with permanent discontinuation, and weighted average daily dose categories. Denominators for calculating percentages will be based on the number of subjects who received at least one dose in the treatment group summarized. All exposure data will be presented in a subject data listing.

10.1.2. Weighted Average Daily Dose (WADD)

Exposure to OCA will also be summarized by weighted average daily dose (WADD) to account for the flexibility of dose adjustments as specified by the protocol. The weighted average daily dose is defined in Section 10.1.1.

Total duration (days) of exposure to OCA will be calculated as follows:

- Last dose date of OCA – first dose date of OCA +1

The number of days on OCA will be calculated as the cumulative number of days that the subject was actually on OCA. It will be summarized by descriptive statistics as well as by frequency counts and percentages for categories of 3-month intervals.

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

- (# of days consumed OCA during study)/ (# of days planned on OCA, excluding drug holidays) * 100

Descriptive statistics for the total duration of exposure to OCA, number of days on OCA, percent OCA compliance, and weighted average daily dose will be summarized with the following WADD categories as treatment dose groups:

- ≤5 mg
- >5 mg to ≤10 mg
- Total OCA (combining all dose groups).

10.2. Adverse Events (AEs)

All AE summaries will be restricted to treatment-emergent adverse events (TEAEs), which are defined as any AEs that newly appear, increase in frequency, or worsen in severity following initiation of study medication. If it cannot be determined whether the AE is treatment emergent due to a partial onset date, then it will be counted as treatment emergent. Verbatim terms on eCRFs will be mapped to preferred terms and system organ classes using MedDRA (version 17.1).

Summaries that are displayed by system organ class and preferred terms will be ordered by descending order of incidence for all treated patients of system organ class and by preferred terms within each system organ class. Summaries of the following types will be presented by treatment group:

- Overall summary of TEAEs
- Subject incidence of TEAEs and the total number of entries by MedDRA system organ class and preferred term.
- Subject incidence of drug-related TEAEs by MedDRA system organ class and preferred term. Related AEs are those with relationships reported as "Definite," "Probable," "Possible," or with a missing relationship.
- Subject incidence of serious adverse events (SAEs) by MedDRA system organ class and preferred term.
- Subject incidence of TEAEs leading to investigational product withdrawal or study discontinuation by MedDRA system organ class and preferred term. This is a subset

of the AEs where Action Taken with Study Medication is checked as "Drug Withdrawn" or Subject Discontinued from Study is checked.

- Subject incidence of TEAEs leading to investigational product withdrawal by MedDRA system organ class and preferred term. This is a subset of the AEs where Action Taken with Study Medication is checked as "Drug Withdrawn".
- Subject incidence of TEAEs leading to study discontinuation by MedDRA system organ class and preferred term. This is a subset of the AEs where Subject Discontinued from Study is checked.
- Subject incidence of TEAEs by MedDRA system organ class, preferred term, and maximum severity will be presented. At each level of subject summarization, a subject is classified according to the maximum severity if the subject reported one or more events. AEs with missing severity will be considered severe for this summary.
- Subject incidence of follow-up TEAEs and the total number of entries by MedDRA system organ class and preferred term may be presented. Follow-up TEAEs are defined as those TEAEs with onset after the End of Treatment study visit.

The summary of subject incidence of TEAEs and the total number of entries will be repeated presenting only preferred term ordered by descending order of incidence for all OCA treated patients, excluding system organ class.

The following listings will be presented by subject:

- All AEs
- Serious AEs (subset of the AEs where serious is marked as "Yes")
- Death information will be provided in a separate listing, should any deaths occur
- Severe AEs (subset of AEs where severity is marked as "Severe" or severity is missing)
- Related AEs (subset of AEs where relationship to study medication is marked as "Definite", "Possible" or "Probable")
- AEs leading to withdrawal of investigational product (subset of AEs where action taken with study medication is marked as "Drug Withdrawn")
- AEs leading to Study Discontinuation (subset of AEs where subject discontinued from study is checked)
- Follow-Up AEs with onset after End of Treatment visit.

10.3. Adverse Events of Special Interest

Adverse events of special interest (AESI) are pruritus, and dyslipidaemia. For each of these AESI, subject incidence of TEAEs and the total number of entries by MedDRA system organ class and preferred term will be presented.

10.3.1. Pruritus

Treatment-emergent pruritus, defined as any preferred term including "Prur," will be summarized separately by the MedDRA system organ class (SOC), treatment group, and preferred term (PT) as a subset of all TEAEs.

10.3.2. Dyslipidaemia

AE lipid profile changes, defined in the Dyslipidaemia SMQ, will be reported. See Appendix B for a list of the preferred terms and their associated codes. These events will be summarized separately by the MedDRA system organ class (SOC), treatment group, and preferred term (PT) as a subset of all TEAEs.

10.4. Clinical Laboratory Evaluations

Clinical laboratory evaluations are assessed at the central laboratory. A listing of available laboratory reference/normal ranges for each laboratory parameter will be provided including age, sex, values with units.

Quantitative hematology, coagulation and serum chemistry laboratory parameters will be summarized by treatment group in SI units using arithmetic summary statistics as described in Section 6.2.1 at DB Baseline and at each scheduled post-Baseline visit. Change from DB Baseline will also be summarized in the same manner. Re-tests or unscheduled visit results after first dose of investigational product in the DB phase will not be summarized but will be included in the data listings.

In addition, shift tables (ie, low-normal-high at DB Baseline versus low-normal-high at post-Baseline Visit in a 3-by-3 contingency table) from Baseline to worst value, last value, and at each scheduled post-Baseline visit will be provided for hematology, coagulation, and serum chemistry by treatment group.

Urinalysis results will not be summarized but will be provided in a data listing.

Listings of all laboratory values will flag values outside of the normal range as high (H) or low (L) and indicate whether or not a value is clinically significant (CS), based on investigator judgment.

For laboratory test results that are below the quantifiable limits:

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

For laboratory test results that are above the quantifiable limits:

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

10.5. Vital Signs

Vital signs (oral temperature, sitting heart rate, respiratory rate, and sitting blood pressure [systolic and diastolic]) will be summarized using arithmetic summary statistics as described in Section 6.2.1 at DB Baseline and at each post-Baseline visit. Change from DB Baseline will also be summarized in the same manner. Re-tests or unscheduled visit results prior to the

administration time of the first dose in the DB phase will be included in the calculation of Baseline. Re-tests or unscheduled visit results after the first DB investigational product administration date will not be summarized but will be included in the data listings.

10.6. Electrocardiograms (ECGs)

Overall interpretation results for ECGs and the Investigator interpretation results are collected as normal, abnormal-not clinically significant, and abnormal-clinically significant. Summary of interpretation shifts from DB Baseline to any post-Baseline visit will be presented by treatment group and visit. In addition, results will be listed separately including description of the abnormality and any associated comments.

11. REFERENCES

Elman S, Hynan L, Gabriel V, et al. The 5-D itch scale: a new measure of pruritus. British Journal of Dermatology. 2010;162:587-593.