DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

CB8025-32048 (RESPONSE): A PLACEBO-CONTROLLED, RANDOMIZED, PHASE 3 STUDY TO EVALUATE THE EFFICACY AND SAFETY OF SELADELPAR IN PATIENTS WITH PRIMARY BILIARY CHOLANGITIS (PBC) AND AN INADEQUATE RESPONSE TO OR AN INTOLERANCE TO URSODEOXYCHOLIC ACID (UDCA)

Statistical Analysis Plan

DRAFT: VERSION 1.0
DATE OF PLAN:

28-AUG-2023

STUDY DRUG:
Seladelpar

PREPARED FOR:

CymaBay Therapeutics, Inc.

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

Approval Signature Page: CymaBay Therapeutics, Inc.

DocuSigned by:
PPD
Signer Name: PPD
Signing Reason: I approve this document
Signing Time: 30-Aug-2023 | 10:40 PDT
94DC2290117A4E4682451FB4C024259B

30-Aug-2023 | 10:41 PDT

PPD
Vice President of Biometrics

Date

DocuSigned by:
PPD
Signer Name: PPD
Signing Reason: I approve this document
Signing Time: 30-Aug-2023 | 11:01 PDT
6DF1620E6D814F30826A766013F2F158

30-Aug-2023 | 11:01 PDT

PPD
Vice President of Clinical Development

Date

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

TABLE OF CONTENTS

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

LIST OF TABLES

LIST OF FIGURES

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

ABBREVIATIONS

AE	Adverse event
ALT	Alanine aminotransferase
AMA	Antimitochondrial antibodies
ANA	Antinuclear antibodies
ANCOVA	Analysis of covariance
ALP	Alkaline phosphatase
APAC	Asia-Pacific
AST	Aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
BA	Bile acids
BMI	Body Mass Index
BUN	Blood urea nitrogen
C4	7α-hydroxy-4-cholesten-3-one
CDCA	Chenodeoxycholic acid
CDF	Cumulative distribution function
CERC	Clinical event review committee
CI	Confidence interval
CK	Creatine kinase
CMH	Cochran-Mantel-Haenszel
CRF	Case Report Form
CTMS	Clinical Trial Management System
CTCAE	Common Terminology Criteria for Adverse Events
CSR	Clinical study report
DSMB	Data Safety Monitoring Board
ECG	Electrocardiogram
eDISH	evaluation of Drug-Induced Serious Hepatotoxicity
ELF	Enhanced Liver Fibrosis
ET	Early Termination
FGF19	Fibroblast growth factor 19
FGF21	Fibroblast growth factor 21
GGT	Gamma-glutamyl transferase
HA	Hyaluronic acid
HCV	Hepatitis C Virus
HDL-C	High density lipoprotein cholesterol
hs-CRP	High sensitivity C-reactive protein
IgM	Immunoglobulin M
INR	International normalized ratio
IP	Investigational Product
ITT	Intent-to-treat
IWRS	Interactive web response system
LDH	Lactate dehydrogenase
LDL-C	Low density lipoprotein cholesterol
IL-31	Interleukin-31
LLN	Lower limit of normal
LoV	Last observed value

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

LS means	Least squares means
MedDRA	Medical Dictionary for Medical Affairs
MELD	Model for End-Stage Liver Disease
mITTb	Modified intent-to-treat biopsy
MMRM	Mixed-Effect Model Repeated Measure
MSPN	Moderate to severe pruritus numerical rating scale
NCI	National Cancer Institute
OCA	Obeticholic acid
PD	Pharmacodynamics
PGI-C	Patient Global Impression of Change
PGI-S	Patient Global Impression of Severity
PRC	Pathology review committee
PK	Pharmacokinetics
PP	Per protocol
PT	Preferred Term
PPAR	Peroxisome Proliferator-Activated Receptor
Pro-C3	N-terminal type III collagen propeptide
QoL	Quality of life
RBC	Red blood cells
SAE	Serious adverse events
SAP	Statistical analysis plan
SOC	System Organ Class
SD	Standard deviation
SE	Standard error
TB	Total bilirubin
TC	Total cholesterol
TEAE	Treatment emergent adverse event
TG	Triglycerides
UDCA	Ursodeoxycholic acid
ULN	Upper limit of normal
UNS	Unscheduled
WBC	White blood cells
WHODD	World Health Organization Drug Dictionary

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

1. INTRODUCTION

The statistical analysis plan (SAP) details the planned analyses study as described in protocol CB8025-32048 (RESPONSE): A Placebo-controlled, Randomized, Phase 3 Study to Evaluate the Efficacy and Safety of Seladelpar in Patients with Primary Biliary Cholangitis (PBC) and an Inadequate Response to or an Intolerance to Ursodeoxycholic Acid (UDCA). The content of this SAP is based on the protocol version 4.0, dated 09 February 2022.

2. STUDY OBJECTIVES AND ENDPOINTS

Objectives	Endpoints
Primary	
• **Efficacy:** To evaluate the treatment effect of seladelpar on composite biochemical improvement in cholestasis markers based on alkaline phosphatase (ALP) and total bilirubin at 12 months of treatment compared to placebo	1. Proportion of subjects who are considered responders at 12 months based on the following composite endpoint of ALP and total bilirubin at 12 months requiring a. ALP < 1.67× ULN b. ≥ 15% decrease in ALP c. Total bilirubin ≤ 1.0× ULN
• **Safety:** To evaluate the safety of seladelpar over 12 months of treatment compared to placebo	2. Assessment of treatment-emergent adverse events (TEAEs) (National Cancer Institute [NCI] Common Terminology Criteria for Adverse Events [CTCAE] Version 5.0), biochemistry, and hematology
Secondary	
Key Secondary: • To evaluate the effect of seladelpar on the normalization of ALP values at 12 months of treatment compared to placebo	Key Secondary: 1. Proportion of subjects with ALP ≤ 1.0× ULN at 12 months (ie, normalization)
• To evaluate the effect of seladelpar on pruritus at 6 months of treatment compared to placebo in subjects with baseline moderate to severe pruritus	2. Change from baseline in weekly averaged pruritus NRS at 6 months in subjects with baseline NRS ≥ 4

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

CCI	

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

3.1. Study Design and Population

This is a Phase 3, international, multicenter evaluation of seladelpar in a randomized, double-blind, placebo-controlled, parallel-group study, with seladelpar administered for up to 12 months as a daily oral capsule in subjects with PBC. Approximately 180 subjects were planned to be randomized across approximately 180 sites worldwide.

To enroll, subjects must have confirmed PBC as defined by having any 2 of the following 3 diagnostic criteria at screening: (1) history of ALP above 1.0× ULN for at least 6 months;

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

(2) positive AMA titers (> 1:40 on immunofluorescence or M2 positive by ELISA) or positive PBC-specific ANAs; and (3) documented liver biopsy results consistent with PBC. Please refer to the protocol for detailed inclusion and exclusion criteria.

This study is a pivotal study to evaluate the safety of seladelpar and its effect on cholestasis markers and pruritus in subjects with an inadequate response to UDCA or an intolerance to UDCA. Enrolled subjects must have received UDCA for at least 12 months (> 3 months of stable dose) or have intolerance to UDCA (no UDCA dose in the past > 3 months). During the study, the study drug will be administered as an add-on to the standard-of-care UDCA therapy for subjects who tolerate UDCA; for subjects with UDCA intolerance, the study drug will be administered as a monotherapy.

Subjects are randomly assigned to receive seladelpar 10 mg or placebo in a 2:1 ratio. Subjects will be stratified by ALP level (< 350 U/L versus ≥ 350 U/L) and the presence of clinically important pruritus NRS (< 4 versus ≥ 4) to ensure even distribution across treatment groups. Subjects are considered as formally enrolled in the study at the time of randomization.

Study drug (placebo or seladelpar) is taken in a blinded manner, orally, once daily, for a period of up to 12 months. Subjects receiving UDCA are to continue on UDCA throughout the study.

Subjects are asked to use an electronic diary (e-diary) to evaluate pruritus and QoL during the study participation. E-diaries are dispensed at the Run-in Visit and will include the following questionnaires: pruritus NRS, 5-Dimension (5-D) Itch, PGI-S, PGI-C, and PBC-40. Subjects perform an evaluation of their pruritus on a daily basis via pruritus NRS starting from the Run-in Visit through the first 6 months of treatment. After 6 months, pruritus is evaluated on a monthly basis until the Month 12/End of Treatment (EOT) Visit using pruritus NRS for 7 consecutive days each month. 5-D Itch is evaluated biweekly from the Run-in Visit for the first 6 months of treatment and on a monthly basis after that until the Month 12/EOT Visit. The PGI-S and PBC-40 will start at Run-In and will continue at Randomization, Month 1, Month 3, and then every 3 months through Month 12/Early Termination. PGI-C will start at Month 1 and will continue at Month 3 and then every 3 months through Month 12/Early Termination.

The total duration of participation in the study for each subject is up to ~14 months including the Screening Period (up to 3 weeks), Run-in Period (2 weeks), and Treatment Period (up to 12 months). During the Treatment Period, subjects are seen every 3 months, except for the first on-treatment visit, which is performed 1 month after the initiation of the study drug. After the completion of the Treatment Period, the subjects are invited to enroll into open-label, long-term study CB8025-31731-RE, in which all subjects receive seladelpar. Subjects who decline or are not eligible for CB8025-31731-RE will have a safety follow up visit performed 2 weeks (14 days +3) after the last dose of study drug. Subjects who discontinue study drug treatment for any reason other than a defined PBC clinical outcome in study CB8025-32048 will be asked to stay in the study without the study drug intake.

Subjects who discontinue study drug treatment anytime and do not stay in the study are terminated from the study and complete an Early Termination Visit. For subjects who decline to stay in the study without study drug intake, or who do not participate in study CB8025-31731-RE, a phone call will be performed for PBC outcomes on an annual basis until the study last

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

subject last visit. If a subject terminates study participation at any point after Day 1, an Early Termination Visit is to be completed.

Study visits may occur at the study sites, with the assistance of a home health service, or using virtual technologies based on the sites' determination.

In order to establish liver histology before and after treatment, all subjects are encouraged to have a liver biopsy during the Screening Period (unless a historical biopsy meeting quality standards can be supplied) to evaluate PBC stage at baseline and histology changes after 1 year of treatment from baseline. Subjects willing to undergo this procedure have liver biopsy performed during the Screening Period. The follow-up liver biopsy is to be performed at the Month 12 Visit or at Early Termination if the subject withdraws from the study, provided they have had at least 6 months of treatment. A follow-up liver biopsy is performed only in subjects with baseline liver biopsy. Histology will be evaluated by a pathology review committee (PRC) as specified in a histopathology plan, outlined separately from this SAP.

Subjects are asked to participate in a pharmacokinetic (PK) sample collection to evaluate plasma concentrations of seladelpar and its metabolites. Subjects who consented to participate in this PK sample collection provide a predose sample (-30 minutes prior to dosing) and 2 postdose samples (1 hour ± 30 minutes and at 3 hours ± 30 minutes) at Month 3 and at Month 12 per the assigned schedule. Results will be included in population PK and exposure-response modeling, outlined separately from this SAP.

The CB8025-32048 protocol outlines safety monitoring criteria for subjects with potential drug-induced liver injury, muscle injury, renal injury, and acute pancreatitis, with guidance to the Investigator on actions including stopping study drug, interrupting study drug, down-titrating study drug, and investigating the case.

During the study, subjects will be evaluated for PBC progression by collecting information about PBC clinical outcomes. All PBC clinical outcomes will be adjudicated by a critical event review committee (CERC). The CERC also adjudicates DILI events and the relationship of these events to study drug.

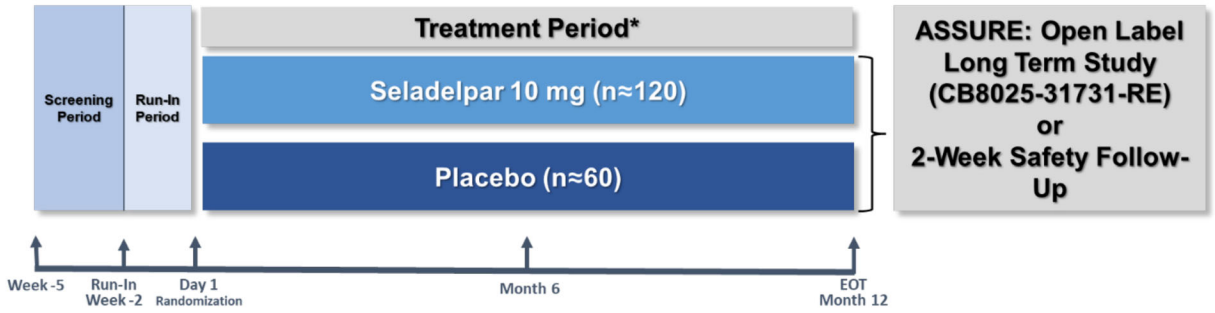
In addition to the CERC, an independent data safety monitoring board (DSMB) was convened to review the study data on a regular basis during the study conduct to ensure subjects' welfare and preserve study integrity.

The primary efficacy analysis is a responder analysis of a composite of biochemical measures after 12 months of treatment with study drug.

See Figure 1 for the study schematic.

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

Figure 1: Study Schematic



Abbreviations: ALP= alkaline phosphatase; ALT= alanine aminotransferase; AST=aspartate aminotransferase; EOT=End of Treatment; UDCA=ursodeoxycholic acid; ULN=upper limit of normal.
* Seladelpar is an add-on to UDCA for subjects with an inadequate response to UDCA in the prior 12 months, or used as a monotherapy in subjects with intolerance to UDCA (last dose of UDCA > 3 months prior to Screening).

3.2. Randomization and Blinding

Subjects are randomized to receive seladelpar 10 mg or placebo in a 2:1 ratio (approximately 120 subjects in the seladelpar 10 mg group and 60 subjects in the placebo group) and are stratified by ALP level (< 350 U/L versus ≥ 350 U/L) and the presence of clinically important pruritus NRS (< 4 versus ≥ 4).

The randomization procedure is performed centrally via an interactive web response system (IWRS) at the Day 1 Visit. The randomization schedule is prepared by an unblinded statistician, separate from the study team.

This is a randomized, double-blind study. During the period from randomization until database lock, the Sponsor study team members responsible for study oversight, subjects, Investigators, and all study-site personnel are blinded to treatment assignment. A DSMB is allowed to review unblinded data as stated in the DSMB Charter. Criteria for emergency unblinding by the PI are outlined in the protocol. If undertaken, emergency unblinding is required to be clearly justified and explained by a comment in the source documentation, along with the date, on which the code is broken, and the identity of the person authorizing the unblinding. The date and time when the Investigator removed the study blind for an individual subject is documented by the IWRS, and an automated blinded notification of the incident will be sent to the Sponsor. The contract research organization's pharmacovigilance team may also be required to break the blind for regulatory reporting purposes.

3.3. Sample Size Considerations

For the purpose of sample size estimation, the placebo group response rate for the primary efficacy endpoint (a composite endpoint of ALP and total bilirubin) evaluated at 12 months is estimated as 20%. The seladelpar 10 mg dose group response rate is estimated as 55%. With the use of a 2-sided test of equality of binomial proportions based on Fisher's exact test at the 0.05 level of significance, a sample size of 180 randomized subjects who received study drug will provide > 90% power to detect a difference between the 10 mg seladelpar group and the placebo group, where any subject who does not provide a 12 month assessment will be considered a non-responder.

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

The key secondary efficacy analysis of normalization of ALP is estimated to have a placebo response rate and a seladelpar response rate of 2.5% and 25.5%, respectively. A sample size of 180 randomized subjects who received study drug will provide > 90% power to detect a difference between the seladelpar and placebo groups, based on a 2-sided Fisher's exact test at a 0.05 level of significance, where any subject who does not provide a 12-month assessment will be considered a non-responder.

The key secondary efficacy analysis of changes from baseline in weekly averaged pruritus NRS at Month 6 sample size calculation is based on a 2-sample 2-sided t-test with a significance level of 0.05. The common standard deviation is estimated as 2. Under these assumptions, a total of 48 randomized subjects who received study drug with a baseline NRS ≥ 4 and NRS at Month 6, provide > 80% power to detect a treatment difference of ≥ 2 between the 10 mg seladelpar and placebo groups.

The assumptions for these power calculations are based on previous results from study CB8025-31735. Additionally, for responder analyses, a dropout rate of approximately 10% was assumed.

3.4. Study Review Boards and Committees

An independent DSMB was established in order to protect subjects' welfare, preserve study integrity, and provide recommendations as needed regarding study conduct. The DSMB is composed of 2 external liver disease clinical experts and a biostatistician. The DSMB meets at predefined timepoints; as-needed meetings based on enrollment, treatment milestones, and safety are also allowed. The DSMB reviews all SAEs, liver-related safety events, and elevations in ALT/AST, serum creatinine, CK, and lipase meeting safety monitoring criteria. Formal minutes and recommendations are provided by the DSMB to the Sponsor regarding additional data requests and the continuation of study conduct as outlined in the current protocol. The DSMB operates under the guidance of an agreed-upon charter. Details of the format, content, and frequency of these meetings are described in the DSMB charter.

A CERC was established to act as an expert, independent panel to review fatal events, hepatic events, and disease progression occurring during the course of the study (per protocol Section 11) to determine if events meet the prespecified event definitions. The members of the CERC include 3 independent hepatologists and/or gastroenterologists. A CERC charter was developed to establish predefined critical event definitions and procedures that will be followed. The CERC members evaluate whether critical event definitions are met during independent voting, remaining blinded to treatment assignment.

The evaluation of liver biopsy tissue from subjects with PBC is intended to provide an important component of a comprehensive safety dataset on seladelpar in its PBC development program. A PRC was formed, consisting of 3 experienced hepatopathologists who serve as the blinded pathologists to read biopsies collected during the study. The PRC is responsible for selecting the scoring system, approving the biopsy reading process, and establishing a framework for evaluating biopsies based on scoring that may guide selection of cases for clinico-pathology adjudication, if necessary. Collectively, these plans are outlined in a histopathology plan and clinico-pathology plan. A clinico-pathology review committee (CPRC) was convened to

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

adjudicate any histopathological cases deemed concerning as possible DILI by the PRC per the histopathology plan.

3.5. Interim Analysis

There were no interim analyses planned for this study. As discussed in Section 3.4, a DSMB was convened to review study safety as described in the DSMB charter.

3.6. Timing of Analyses

The final analysis will take place after last subject last visit has occurred and the database has been locked.

4. DATA ANALYSIS CONSIDERATIONS

All analyses will be conducted based on SAS 9.4 or higher.

All data in the database supporting the objectives of the study will be presented in by-subject data listings.

Unless otherwise stated, all listings will be sorted by treatment group, subject number, visit, and assessment date and time, if appropriate and available.

Unless stated otherwise, continuous data will be summarized by treatment group based on n, mean, median, standard deviation (SD), first quartile (Q1), third quartile (Q3), minimum value, and maximum value.

Unless stated otherwise, categorical data will be summarized by treatment group using n and percentage based on the number of nonmissing values.

- The number of missing values will be presented as a separate category with no percentage, but only if one or more subjects are missing data.
- Counts of zero will be presented without percentages.

Precision

- Mean, Median, Q1, and Q3: one additional decimal place to that reported for Minimum and Maximum
- SD: two additional decimal places than the Minimum and Maximum
- Percentages: reported to one decimal place
- P-values will be reported to four decimal places. If the value is below 0.0001, it will be noted as < 0.0001; if the value is above 0.9999, it will be noted as > 0.9999.

Unless otherwise noted, statistical inference will be based on a 2-sided 0.05 significance level, and 2-sided 95% confidence intervals will be produced.

All data up to the time of study completion/withdrawal from the study will be included in the analysis, regardless of duration of treatment.

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

Numbering for data displays will be based on ICH E3.

4.1. Stratification and Covariates

All statistical tests, unless stated otherwise, will be stratified by baseline levels described in Section 4.2.

4.2. Evaluation of Subgroups

The following subgroup analyses will be conducted, as appropriate, as described in Section 9 (for efficacy analyses) and Section 11 (for safety analyses). A minimum of 5 subjects in each treatment group is required to conduct subgroup analyses.

- Age categories (age at screening: < 65, ≥ 65 years; age at PBC diagnosis < 50, ≥ 50 years)
- Sex (Female, Male)
- Race (White, Black, Asian, Other)
- Region (North America, Europe, Rest-of-World)
- Baseline ALP (< 350 U/L, ≥ 350 U/L)
- Total bilirubin (TB) (< 0.6 × ULN, ≥ 0.6 × ULN)
- Pruritus NRS (< 4, ≥ 4)
- UDCA use vs UDCA intolerance
- Prior use of OCA and/or fibrates (yes, no)
- Cirrhosis (yes, no)
- TB (≤ 1 × ULN, > 1 × ULN)

The consistency of the treatment effects across stratification levels will be assessed using the same approach as specified for the primary efficacy analysis. Subgroup analyses for ALP will not be stratified or adjusted by ALP stratum; subgroup analysis for pruritus NRS will be performed similarly.

4.3. Multiple Comparisons and Multiplicity

Study-wide Type I error will be maintained at 5% using a hierarchical fixed-sequence methodology for the primary and key secondary efficacy analyses. The fixed-sequence approach for the primary and key secondary analyses is as follows:

- If the primary efficacy analysis is positive for seladelpar 10 mg versus placebo at a 2-sided 0.05 significance level, then the 2 key secondary endpoints will be analyzed hierarchically in the following order:
    - Normalization of ALP at Month 12 (seladelpar 10 mg versus placebo): If negative at a 2-sided 0.05 significance level, no further inferential testing will be performed. Otherwise, if positive, then testing will proceed.
    - Changes from baseline to Month 6 in pruritus NRS (seladelpar 10 mg versus placebo) will be evaluated at a 2-sided 0.05 significance level.

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

5. GENERAL DATA HANDLING CONVENTIONS

5.1. Assigned and Actual Treatment

Assigned treatment groups (seladelpar 10 mg and placebo) will be determined based on IWRS assignments. Actual treatment group will be determined based on whether a subject received seladelpar during the study; subjects who take any amount of seladelpar will be reported in the seladelpar 10 mg actual treatment group. Subjects who did not take seladelpar during the study will belong to the placebo actual treatment group.

5.2. Study Drug

Study drug refers to either seladelpar or placebo.

5.3. Study Day and Duration Variables

Unless specified otherwise, reference date calculations will be defined as the following:

- date of interest – reference date + 1 when the date of interest ≥ reference date;
- otherwise, date of interest – reference date.

If either date is missing, reference date calculations will not be performed. Date imputation will be performed as identified in Section 5.7.

For instance, study day will be based on the treatment start date as the reference and would either have a negative value if collected before the first dose date or a positive value if collected on or after the day of drug dosing. There will be no study day zero.

Duration on study is defined as the end of study date (collected on Study Termination CRF page) – treatment start date + 1. Duration of treatment is defined as treatment end date – treatment start date + 1, where treatment end date is the date of last dose of study drug for this protocol.

5.4. Reference Dates

- Screening date is defined as the date on which a subject was screened for trial entry, as recorded in the eCRF.
- Randomization date is defined as the date on which the subject is randomized to study treatment.
- Treatment start date is defined as the date of first dose of study drug.
- Treatment end date is defined as the date of last dose of study drug for this protocol.
- The calculation of age will use the screening date as its reference date.
- Safety data, such as AEs and laboratory assessments, will use the treatment start date as a reference date.
- Efficacy will use the treatment start date as a reference date.

5.5. Study Time Periods

The screening period is up to 3 weeks, starting from the date of signing the informed consent and ending prior to initiation of the Run-in period. The Run-in period occurs 2 weeks prior to a subject's treatment start date.

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

Day 1 begins the Treatment Period and is the date against which all subsequent visits will be timed except for the safety follow-up period. The Treatment Period concludes on the date of last dose. After the completion of the Treatment Period, subjects will be invited to enroll into open-label, long-term study CB8025-31731-RE. Subjects who decline participation in study CB8025-31731-RE will have a safety follow-up visit performed 2 weeks (14 days + 3) after the last dose of the study drug under protocol Version 4.0 (changed from 4 weeks + 7 days in protocol Version 3.0).

The window used for summarizing treatment-emergent adverse events is between a subject's treatment start date through 30 days following the treatment end date.

All AEs collected in the study are presented in a subject listing.

5.6. Baseline, Post-Baseline Changes, and Last Observed Value (LOV)

Unless stated otherwise, baseline values will be based on the last nonmissing assessment evaluated prior to the first administration of study drug.

Baseline for chemistry, hematology, other laboratory quantitative measures, ECGs, vital signs, PBC-40, and 5-D Itch scales, as appropriate, will be calculated as the arithmetic mean of applicable measurements at Screening, Run-in, Day 1, and unscheduled assessments prior to or on Day 1.

Baseline pruritus NRS is defined as the mean of all daily recorded scores during the Run-in Period and on Day 1. The same baseline will also be used for the worst monthly summaries. Worst monthly post-baseline will be the maximum of the weekly NRS averages during the same period in Table 3 and calculated only up to Month 6.

Change from baseline is defined as: value – baseline value.

Percentage change from baseline is defined as: (value – baseline value)/baseline value × 100%.

Most extreme change: The maximum most extreme change will be based on the maximum post-baseline value; the minimum most extreme change will be based on the smallest post-baseline value. This calculation will consider all assessments collected during the treatment-emergent period, scheduled or unscheduled.

The last observed value (LoV) will be the last nonmissing value collected during the treatment-emergent period. This value will consider all assessments, scheduled or unscheduled.

5.7. Imputation of Partial Dates

Adverse Events and Concomitant Medications

- If the AE start date is completely missing, or if the subject was not treated with study drug, no imputation will be conducted. An AE with completely missing start date will be considered as a TEAE unless the AE end date is present and before the first dose date.
- If the AE start date is missing day and month, do the following:
    - If the treatment start date is missing or the AE start year does not fall in the same year as that of the treatment start date or if the AE record contains information to

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

indicate that the event ended before the treatment start date (e.g., the AE end date month and year are earlier than the treatment start date, or the full AE end date is known and occurs earlier than the treatment start date), then set the AE start month and day to January 1st.

- Otherwise, set the AE start date to the treatment start date.

- If only the AE start day is missing, do the following:
  - If the study treatment start date is missing or the AE start month and year does not fall in the same month and year as that of the treatment start date or if the AE contains information to indicate that the event ended before the treatment start date, then set the AE start day to the 1st day of the month of the AE start date.
  - Otherwise, set the AE start date to the treatment start date.

Generally, missing dates will be imputed conservatively in such a way that the duration of the adverse event is considered with the longest possible duration and such that, whenever the adverse event may potentially start after first dose administration, the adverse event will be handled as a TEAE.

These imputation rules will also be applied to concomitant medication dates.

5.8. Multiple Assessments and Visit Windows

Unless specified otherwise, the following rules are applied to assign analysis visits to assessments obtained at scheduled and unscheduled protocol visits.

Unscheduled data may be included in summaries of most extreme, baseline, and LoV values; summaries of specific abnormalities any time post-baseline; and subject data listings.

Generally, analysis visits may include data up to seven days after Treatment End Date, except Safety Follow-up visit.

5.8.1. General Visit Windows

The visit windows and the target days for each analysis visit are listed Table 1.

Table 1: General Visit Windows for Data Presentations

Scheduled Visit	Target Day[1]	Study Day Analysis Visit Window Range
Month 1 (Week 4)	Day 29	Days 2 - 56
Month 3 (Week 12)	Day 85	Days 57 - 133
Month 6 (Week 26)	Day 183	Days 134 - 227
Month 9 (Week 39)	Day 274	Days 228 - 318
Month 12 (Week 52)	Day 365	Days 319 - Treatment End Date + 7
Safety Follow-up (14 + 3 days after EOT)[2]	Day 14 after last dose	≥ Treatment End Date + 8

Abbreviations: EOT=end of treatment.
[1] Study days are relative to Day 1 (the treatment start date).

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

[2] The safety Follow-up visit will not be included in inferential modeling and applies to subjects who are not enrolled into the long-term study (CB8025-31731-RE).

If a subject has 2 or more scheduled visit assessments in one visit window, the scheduled visit assessment closest to the target day will be used as the analysis visit assessment for that visit window. If 2 scheduled visit assessments are equidistant from the target day within a visit window, the later scheduled visit assessment is used. If a visit window has no scheduled visit assessments but does have at least one unscheduled visit assessment, then the unscheduled visit assessment closest to the target day will be used.

5.8.2. NRS Pruritus Data Windows

Pruritus NRS at Month 6 is a key secondary efficacy endpoint, but data collected at Month 1, 3, 6, 9, and 12 will also be summarized and used in MMRM analyses as described in Section 9.1.2. Weekly pruritus score at these timepoints through Month 6 will be calculated by recording the daily pruritus score each day for seven days and then taking the mean value of the seven days' daily recorded data. If any data are available for a given week, the available NRS results will be used for the calculation of the weekly mean. The analysis windows described in Table 2 will be applied to calculate weekly means planned for summarization:

Table 2: Weekly NRS Windows for Data Summarization

Scheduled Visit	Target Day	Study Days Used for Weekly Mean
Month 1 (Week 4)	Day 29	Days 23 - 29
Month 3 (Week 12)	Day 85	Days 79 - 85
Month 6 (Week 26)	Day 183	Days 177 - 183
Month 9 (Week 39)	Day 242	Days 228 - 255
Month 12 (Week 52)	Day 326	Days 312 - 378

Study days are relative to Day 1 (the treatment start date).

More generally, weekly means and maximum values of NRS for Week X up to Week 26 will be calculated for data imputation purposes based on a study day window that follows the following pattern: the weekly mean for X will be based on Days (X*7)-5 to X*7+1. For Week 39 and Week 52, the target day and analysis windows are chosen to align with the scheduled collection times of NRS, and a wider range is adopted to accommodate subjects' 7-day data falling into different time intervals around the scheduled visits. Imputation methods for weekly NRS pruritus data can be found in Section 9.1.3.

As an additional sensitivity analysis, NRS scores will be averaged over months. For the purpose of this analysis, observed data will be averaged within monthly windows as described in Table 3.

Table 3: Monthly NRS Windows for Data Summarization

Scheduled Month	Target Day	Study Days Used for Monthly Mean
Month 1 (Week 4)	Day 29	Days 2 - 29
Month 3 (Week 12)	Day 85	Days 58 - 85
Month 6 (Week 26)	Day 183	Days 156 - 183
Month 9 (Week 39)	Day 242	Days 228 - 255

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

Month 12 (Week 52)	Day 326	Days 312 - 378

Study days are relative to Day 1 (the treatment start date).

Monthly means will be calculated for data imputation purposes based on a study day window that follows the following pattern: the monthly mean for Week X will be based on Days X*7-26 to X*7+1. If any data are available for a given month, the available NRS results will be used for

Handling of missing dates is described in Section 5.7. Otherwise, missing data is not imputed for analysis.

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

6. STUDY SUBJECT DATA

6.1. Analysis Populations/Sets

6.1.1. All Subjects Screened Analysis Set

The All Subjects Screened set is defined as all subjects who have been screened for enrollment in the study regardless of whether they were enrolled in the study. This set will be used for summarizing reasons for screen failures.

6.1.2. Intent-to-treat (ITT) Analysis Set

The Intent-to-treat (ITT) analysis set is defined as any subject who is randomized into the study and receives at least 1 dose of study drug. The ITT analysis set will be the primary analysis set used for efficacy analyses, with the exception of secondary endpoints evaluated for subjects with moderate to severe pruritus. Subjects will be analyzed according to randomized treatment assignment.

6.1.3. Moderate to Severe Pruritus NRS (MSPN) Analysis Set

The moderate to severe pruritus NRS (MSPN) analysis set includes subjects in the ITT analysis set who have a baseline NRS value ≥ 4. The MSPN analysis set will be the primary analysis set for secondary endpoints based on NRS evaluations. Subjects will be analyzed according to randomized treatment assignment.

6.1.4. Per-protocol Analysis Set

The per-protocol (PP) analysis set includes any subject who is in the ITT analysis set, has at least 1 postbaseline ALP and total bilirubin evaluation, and does not have a protocol violation that is deemed to impact the efficacy analysis. Selected analyses of efficacy will be conducted using the PP analysis set. If the PP analysis set differs from the ITT analysis set by less than 5 subjects, then PP analyses will not be performed.

6.1.5. Biopsy Analysis Set

The biopsy analysis set includes any subject who has a baseline or Month 12/ET biopsy (see Section 12). The biopsy analysis set will be used to examine histopathology changes on treatment. The review and analysis of the biopsy tissue samples are described in a separate histopathology plan. The results will be analyzed per the histopathology SAP and presented in a separate report.

For this SAP, the number of subjects who contributed a liver biopsy sample at each timepoint for analysis will be summarized based on treatment received. The timepoints of assessment will be listed.

6.1.6. Safety Analysis Set

The safety analysis set is defined as any subject who receives at least 1 dose of study drug. Subjects will be included in the group based on actual treatment received, should this differ from the treatment assignment. All safety analyses will be completed using the safety analysis set.

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

6.1.7. Pharmacokinetics Analysis Set

The pharmacokinetics (PK) analysis set includes any subject who participates in the PK sample collection. All PK analyses will be completed using the PK analysis set. Future pooling of the concentration data from this study with data from other studies to facilitate development and/or updating of a population PK model will be reported separately. For this SAP, only the sample collection dates / time and concentration results will be listed.

7. SUBJECT DISPOSITION

7.1. Disposition

The number of subjects screened, the number and percentage of those subjects who were screen failures, and reasons for screen failures will be summarized based on the All Subjects Screened analysis set.

For treatment disposition, the number and percentage of subjects who were randomized and treated, completed treatment, discontinued treatment with reasons for treatment discontinuation, completed treatment and completed follow-up period, completed treatment and did not complete follow-up period, agreed to participate in the long-term study CB8025-31731-RE, or declined to participate in CB8025-31731-RE will be presented by treatment group and overall for all randomized subjects.

For study disposition, the number and percentage of subjects who completed study, and those who discontinued study with reasons for discontinuation, will be presented by treatment group and overall for all randomized subjects.

A summary of subjects included in each analysis set will be provided overall and by treatment group for all randomized subjects. Separate listings of subject eligibility, analysis set inclusion, and final subject disposition will be provided. A listing of subjects excluded from the PP analysis set, with reasons for exclusions, will also be provided.

7.2. Protocol Deviations

Protocol deviations will be recorded within a Clinical Trial Management System (CTMS) and undergo cross-functional team review including medical monitors, clinical operations, and biostatisticians prior to database lock.

The number and percentage of subjects with significant protocol or ICH/GCP deviations will be summarized by treatment group for the ITT analysis set. Subjects will be counted once within each deviation category regardless of how many deviations they have in that category.

A comprehensive listing of all subjects with protocol deviations in the ITT analysis set will be provided. COVID-19-related protocol deviations will be provided in a listing.

The Sponsor will review the significant protocol deviations before database lock and unblinding and use medical judgement to select the subjects that had deviations impacting the analysis of

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

efficacy to determine the PP analysis set. The excluded subjects from the PP analysis set will be listed.

8. DEMOGRAPHIC AND BASELINE CHARACTERISTICS

8.1. Demographics

Subject demographics consist of the following:

- Age (years)
- Age at screening (< 65, ≥ 65 years)
- Sex (female, male)
- Race (White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Multiple, Other, Declined to Answer)
- Ethnicity (Hispanic or Latino, not Hispanic or Latino, Declined to Answer)
- Region (North America; Latin America; Europe, the Middle East and Africa; Asia Pacific)
- Weight (kg)
- Height (cm)
- Body mass index (BMI; kg/m^2)

Baseline characteristics include the following:

- Cirrhosis (yes)
  - Child-Pugh Score
  - Child-Pugh Class
- Portal hypertension (Yes or No)
- AMA status (positive, negative, equivocal)
- Rotterdam stage of disease
  - Mild: normal total bilirubin and normal albumin
  - Moderately advanced: abnormal albumin or abnormal total bilirubin
  - Advanced: abnormal albumin and abnormal total bilirubin
- MELD score category (<12, ≥12)
- MELD score
- Liver stiffness (kPa) by FibroScan® (categorized into F0, F1, F2, F3, F4, where the cut-off values for F1, F2, F3, and F4 are 7.1, 8.8, 10.7 and 16.9 kPa, respectively [Corpechot 2012])
- Enhanced liver fibrosis (ELF) score
- Timing of historical liver biopsy performed prior to baseline (this parameter will be listed only, not summarized)
- UDCA intolerance (yes, no)
- Duration of prior UDCA usage (years)
- Total daily UDCA dose (mg and mg/kg)
- Prior use of OCA and/or Fibrates (yes, no)
- ALP concentration (U/L)

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

- ALP level (< 350 U/L and ≥ 350 U/L)
- Total bilirubin concentration (mg/dL)
- Total bilirubin level (≤ 1× ULN, >1 - ≤ 2× ULN, > 2× ULN)
- Total bilirubin level (< 0.6× ULN, ≥ 0.6× ULN)
- Direct bilirubin concentration (mg/dL)
- ALT (U/L)
- AST (U/L)
- GGT (U/L)
- Platelet (10^3/uL)
- International normalized ratio (INR)
- Albumin (g/dL)
- Pruritus NRS (observed value, and categories < 4 and ≥ 4)
- UK-PBC risk score
- GLOBE score
- 5' nucleotidase
- Total cholesterol
- Triglyceride
- HDL-C
- Non-HDL-C (calculated as Total Cholesterol – HDL-C)
- LDL-C

Subject demographics and baseline characteristics will be summarized by treatment group and overall for the ITT, PP (if applicable), and safety analysis sets and will be presented in subject data listings.

8.1.1. Subjects with Cirrhosis

Baseline characteristics for subjects with cirrhosis will be provided and include the following:

Cirrhosis
Rotterdam Stage
MELD Score
CP Score
CP class
    CP-A
    CP-B
Supporting information for the diagnosis of cirrhosis[a]
    Liver biopsy
    Current decompensated liver disease
    Liver stiffness by FibroScan®
    Radiological evidence
    Laboratory finding
    Clinical determination by the Investigator
Portal hypertension - Yes
    Esophageal varices

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

- Ascites
- Splenomegaly[b]
- Non-esophageal varices[b,c]
- Thrombocytopenia[b]

Baseline labs

- Alkaline phosphatase
- Total bilirubin
- Direct bilirubin
- ALT
- AST
- Albumin
- Platelets
- INR

Liver stiffness by transient elastography (FibroScan®)

- Mean +/- SD
- Min, Max

Notes:

[a] Supporting information for the diagnosis of cirrhosis are from subjects' Month 1 liver status evaluation data.

[b] The evidence was identified by a search of the subjects' medical history.

[c] Non-esophageal varices include duodenal varices, gastric varices, intestinal varices, splenorenal shunt, varicose veins of abdominal wall, and/or gastric antral vascular ectasia.

8.2. Medical History

A detailed medical history will be coded by System Organ Class (SOC) and Preferred Term (PT) using Medical Dictionary for Regulatory Activities (MedDRA). The dictionary version used for reporting the study will be described in the relevant table and listing footnotes. The number and percentage of subjects with any medical history will be summarized overall and for each system organ class (SOC) and preferred term (PT) for the ITT analysis set. A summary by PT only for the ITT analysis set will also be provided.

8.3. PBC History

PBC history will be summarized for the ITT, MSPN, PP (if applicable), and safety analysis sets. In addition to the information collected on PBC History Form, duration of PBC (time [in years] from diagnosis date to the informed consent date) will also be summarized.

8.4. Prior and Concomitant Medication

Concomitant medications are defined as any medications taken during the treatment emergent period (refer to Section 5.4). Prior medications are defined as those that were started before the treatment start date regardless of end date.

Medications reported on the Prior and Concomitant medications case report form (CRF) pages will be coded by the World Health Organization Drug Dictionary (WHODD) version March 2021 format B3 and will be summarized by Anatomical Therapeutic Chemical (ATC) coding

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

system class level 2, generic name by treatment group for the safety analysis set. Subjects taking more than one medication in the same generic name or ATC class will be counted once. The dictionary version used for reporting the study will be described in the relevant table and listing footnotes.

For the purpose of inclusion in prior and/or concomitant medication tables, incomplete medication start dates will be imputed based on the method discussed in Section 5.7.

A listing of prior and concomitant medications will be presented.

8.5. Concomitant Procedures

Concomitant procedures are defined as any therapeutic intervention (e.g., surgery/biopsy, physical therapy) or diagnostic assessment (e.g., blood gas measurement, bacterial cultures) performed during the treatment emergent period (refer to Section 5.4). Prior procedures are those that were performed prior to the treatment start date.

The number and percentage of subjects who had any concomitant procedures will be summarized by procedure and treatment group based on the safety analysis set.

A listing of prior and concomitant procedures will be presented.

8.6. Study Drug and UDCA Exposure and Compliance

8.6.1. Study Drug Exposure

Exposure to study drug will be summarized by treatment group using descriptive statistics in the safety analysis set. Exposure (days) is defined as duration of treatment (see Section 5.3). The number and percentage of subjects with the following exposure categories will be summarized:

- 1 to ≤ 4 weeks (1 – 28 days)
- 4 to ≤ 8 weeks (29 – 56 days)
- 8 to ≤ 12 weeks (57 – 84 days)
- 12 to ≤ 26 weeks (85 – 182 days)
- 26 weeks to ≤ 39 weeks (182 – 273 days)
- 39 weeks to ≤ 52 weeks (274 – 364 days)
- 52 weeks (> 364 days)

The total cumulative dose will be defined as the sum of the actual dose received across all study days. Average daily dose is cumulative dose divided by duration of treatment. The average daily dose and cumulative dose will be summarized using descriptive statistics over the Treatment Period, using the formula as shown below:

Total cumulative dose (mg) = sum of {[(dispensed number of capsules – max (returned number of capsules, number of doses missed)) *dispensed dose amount per capsule] for each occurrence of drug dispensed visit}. If no information on returned or missed dose is available, it is assumed that the subject has taken all planned doses.

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

Average daily dose = cumulative dose / (exposure in days).

Listings will include all available exposure data.

8.6.2. UDCA Exposure

Exposure to UDCA will be summarized similar to study drug.

For UDCA, cumulative dose = sum of [dose strength × (dosing duration in days – (missed number of doses)/(dosing frequency)) at each visit]. If the number of UDCA doses missed at a visit is missing, it is assumed that the subject has taken all planned doses.

Listings will include all available exposure data.

8.6.3. Study Drug Compliance

Overall treatment compliance for study drug will be calculated and summarized by treatment group for the safety analysis set. The compliance rate (%) will be calculated for each subject as noted below.

Compliance rate (%) will be calculated as:

$$100\% \times \frac{\sum number\ of\ capsules\ consumed\ at\ each\ visit}{planned\ total\ capsules\ for\ treatment\ duration}$$

At each visit, number of capsules consumed = (number of capsules dispensed) – max (number of capsules returned, number of doses missed). If no information on returned or missed dose is available, it is assumed that the subject has taken all planned doses.

Compliance will be cumulatively summarized using descriptive statistics during 3-, 6-, 9-, and 12-month treatment periods, based on boundaries using the Target Days per Section 5.8. The number and percentage of subjects will be summarized by treatment groups and categories (e.g., < 80%, ≥ 80% to ≤ 120%, and > 120%) for each treatment group.

8.6.4. UDCA Compliance

UDCA compliance will be estimated based on the number of days when doses are taken and will be summarized by treatment group. UDCA compliance will be calculated as:

$$100\% \times \frac{total\ duration\ of UDCA\ exposure\ in\ days - (number\ of\ UDCA\ doses\ missed)/(dosing\ frequency)}{total\ duration\ of\ UDCA\ exposure\ in\ days}$$

If the number of UDCA doses missed at a visit is missing, it is assumed that the subject has taken all planned doses.

A subject data listing of study drug accountability records will be provided.

8.6.5. Drug Interruptions and Dose Adjustments

Subjects who meet the specific safety monitoring criteria as described in the protocol (Section 10) or have tolerability issues may have a dose down-titration. Subjects initially assigned to

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

10 mg will be down-titrated to 5 mg in a blinded manner. Subjects initially assigned to placebo will be down-titrated to remain on placebo in a blinded manner.

Subjects who experience a clinically important AE that, in the investigator's clinical judgment, warrants a dose reduction are also eligible for a similar dose down-titration. Dose down-titration must be approved by the medical monitor and will be performed in a blinded manner.

The number and percentage of subjects who had dose adjustment (down-titration and up-titration) will be summarized by treatment group for the safety analysis set.

Subjects with lab abnormalities will be monitored closely and may interrupt study drug or discontinue study drug if the criteria are met. A detailed description of the safety monitoring and withdrawal criteria including decision rules can be found in Section 10 of the protocol.

The number and percentage of subjects with dose interruptions, number of dose interruptions, and reasons for dose interruptions as identified on the study drug dose interruption CRF page will be summarized and presented for the safety analysis set.

Separate listings of subjects who experienced dose down- titration or interruptions will be provided.

9. EFFICACY

All efficacy analyses will be conducted on the ITT analysis set, with the exception of analyses related to NRS pruritus score data and liver histopathology data. Unless specified otherwise, analyses of NRS data will be conducted on the MSPN analysis set. The primary and key secondary analyses will also be evaluated based on the PP analysis set (if applicable) where specified. Unless specified otherwise, data collected following treatment discontinuation will not be included in data summaries or inferential analyses. All endpoint data will be summarized by visits appropriate to data type based on definitions in Section 5.8.

9.1. Primary Efficacy Endpoint and Analyses

9.1.1. Primary Efficacy Analysis

The primary efficacy analysis for the proportion of subjects who are considered responders for the primary efficacy endpoint will be based on the proportion of subjects achieving the following composite endpoint evaluated at 12 months (Month 12):

- ALP < 1.67× ULN
- ALP decrease of ≥ 15%
- Total bilirubin ≤ 1.0× ULN

Any subject who does not provide an assessment, who has discontinued treatment prior to the specified time point for response evaluation, or who otherwise has missing data will be considered a nonresponder. The incidence of response at Month 12 will be summarized; analyses for the composite endpoint will be completed using a Cochran-Mantel-Haenszel (CMH) test. The

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

CMH analysis will be stratified by the baseline randomization stratum and will be conducted on the ITT analysis set. The risk difference and 95% CI using Miettinen and Nurminen will also be provided. Statistical significance of the difference between placebo and seladelpar will be defined as a 2-sided $p \leq 0.05$.

If assumptions required for the CMH test are not met, a pooling of stratum will be applied. If the Mantel-Fleiss (1980) criterion is less than 5, this will result in stratum removal. The pruritus stratum will be removed first, and the data will be analyzed using a CMH test stratified by baseline ALP randomization level. Furthermore, both pruritus and ALP strata will be removed and, in this case, a Pearson's chi-squared test will be used with a 95% CI (Wald) for differences in response percentages. Further, when using the Pearson chi-squared test, if at least one expected cell count still falls below 5, the Fisher's exact test and exact 95% CI (Santner and Snell, 1980) will instead be used for inferential analyses.

For the CMH test, the Breslow-Day test will be used to check the assumption of homogeneity of treatment effect across stratum. If the assumption fails, the proportion of responders in each treatment along with its corresponding CIs and the comparison between treatments will be reported for each stratum separately.

Subgroup analyses for the primary endpoint will be performed as described in Section 4.2. A forest plot for the subgroup analyses will be provided.

9.1.2. Key Secondary Efficacy Endpoints

Study-wide Type I error will be maintained at 5% using the hierarchical fixed-sequence methodology for the primary and key secondary efficacy analyses as described in Section 4.3.

The key secondary efficacy analysis for the proportion of subjects who achieve normalization of ALP at 12 months (Month 12) will be conducted in the ITT analysis set using the same approach specified for the primary efficacy analysis in Section 9.1.1.

Change from baseline in weekly averaged pruritus NRS at 6 months will be analyzed using a mixed-effect model for repeated measures (MMRM) for subjects in the MSPN analysis set. The model will include terms for baseline NRS, baseline randomization stratum (ALP level < 350 U/L versus ALP level ≥ 350 U/L), treatment group, week, and treatment-by-week interaction. The MMRM will model repeated measures of NRS changes from baseline at the timepoints described in Table 2 up to Month 6. If a timepoint is missing, it will be imputed as an average of the two adjacent weekly averages (at most one week apart); otherwise, it will be imputed by the adjacent weekly average which is present. For example, if a subject in the study is missing Week 23 and Week 24 data, Week 23 would be imputed based on Week 22 average while Week 24 would be imputed based on Week 25 data. Further, a subject who discontinues prior to or during Week 24 would not have an imputed value for Week 26. Data collected after Month 6 will not be used for imputation.

Treatment by baseline NRS interaction will be explored and added as a term if an interaction is noted to be present (p-value < .05). LS means for the NRS change from baseline by treatment and the associated standard errors, the LS means for the difference between treatment groups, and the associated 2-sided 95% CIs and 2-sided p-values, will be derived from the MMRM

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

model. Significance for this key secondary efficacy analysis is based on the treatment difference at Week 26 (i.e. Month 6). An unstructured covariance matrix will be tried first for the MMRM. If the model fails to converge, the following covariance structures will be implemented in order until the model converges: heterogenous Toeplitz, heterogenous compound symmetry, and then compound symmetry. The Kenward-Roger correction for the denominator degrees of freedom will be applied.

Observed values and changes from baseline in weekly averaged pruritus NRS will be summarized at baseline and by each post-baseline week, as appropriate, in Table 2. Imputation of missing weekly averages will be as described for the MMRM model. Data collected after Month 6 will not be used for imputation, nor will visits after Month 6 be imputed.

Subgroup analyses for the key secondary endpoints will be performed as described in Section 4.2 if applicable. A forest plot for the subgroup analyses will be provided.

9.1.3. Sensitivity Analyses for Primary and Key Secondary Analyses

The primary and key secondary analyses will be repeated for the PP analysis set (if applicable; see the requirement in Section 6.1.4) as an initial sensitivity analysis. NRS analysis will be based on the intersection of MSPN and PP analysis sets.

Change from baseline in pruritus NRS will additionally be analyzed based on the monthly averages based on the windowing described in Section 5.8.2. Another analysis of change from baseline in pruritus NRS will be based on an MMRM including all post-baseline weekly averages for the model described in Section 9.1.2. Testing for this model will be evaluated based on a contrast of averaged weekly scores from Week 14 through Week 26.

Additional sensitivity analyses will be conducted as follows.

Complete Case Analysis

The primary and secondary efficacy endpoints will be evaluated as described in Sections 9.1.1 and 9.1.2 based on a complete case analysis for the ITT analysis set or MSPN for NRS analysis. No data will be imputed for these analyses (i.e., subjects with missing data will be excluded and nonresponse will not be assumed for missing values).

Treatment Policy Strategy Analysis

The primary and secondary efficacy endpoints will be evaluated as described in Sections 9.1.1 and 9.1.2 using the ITT analysis set or MSPN for NRS analysis and all data collected regardless of treatment status if applicable. Data collected following treatment discontinuation will be included in data summaries or inferential analyses. For the composite endpoint and ALP normalization, Month 12 (Week 52) from Table 1 will no longer cut off at treatment end date, but will include Days 319 to 378 regardless of treatment end date. For NRS analysis, data collected after treatment end date that are previously discarded in the primary analysis will be included in the treatment policy strategy analysis.

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

Control-based Multiple Imputation

A control-based multiple imputation method (Ratitch and O'Kelly 2011) will be performed to assess the robustness of the primary analysis. The imputation method serves as an "as treated" model, assuming the placebo group data are adequate to impute missing seladelpar treatment group values after drug discontinuation. This method also uses placebo group data to impute missing placebo group values after drug discontinuation.

In this sensitivity analysis, intermittent missing values will be imputed by visit (or based on weekly averages, as appropriate) using a Markov Chain Monte Carlo (MCMC) methodology which assumes a multivariate normal distribution over all variables included in the imputation model. Intermittent missing data will first be multiply imputed, fitting a monotone pattern based on an MAR model, using SAS MI by treatment group. The remaining monotone missing values for the seladelpar and placebo arms will then be multiply imputed using the placebo group profile (i.e., a copy-reference imputation) using PROC MI and the MNAR statement.

This method will be applied to laboratory data to impute ALP and bilirubin data at visits as described in Section 5.8 for the ITT analysis set. Month 12 primary endpoint and ALP normalization proportions will then be analyzed using the method described in Section 9.1.1. Following this, CMH test results for each imputation will be combined using Rubin's combination rules to compute the p-value based on the Wilson-Hilferty transformation (Ratitch 2013).

For the analysis of weekly NRS averages, datasets will be multiply imputed using all valid weekly averaged data from the visits in Table 2 for the MSPN analysis set. The multiply imputed datasets will each be analyzed based on the MMRM analysis model specified in Section 9.1.2; final, combined inference will be based on Rubin's combination rules (SAS MIANALYZE).

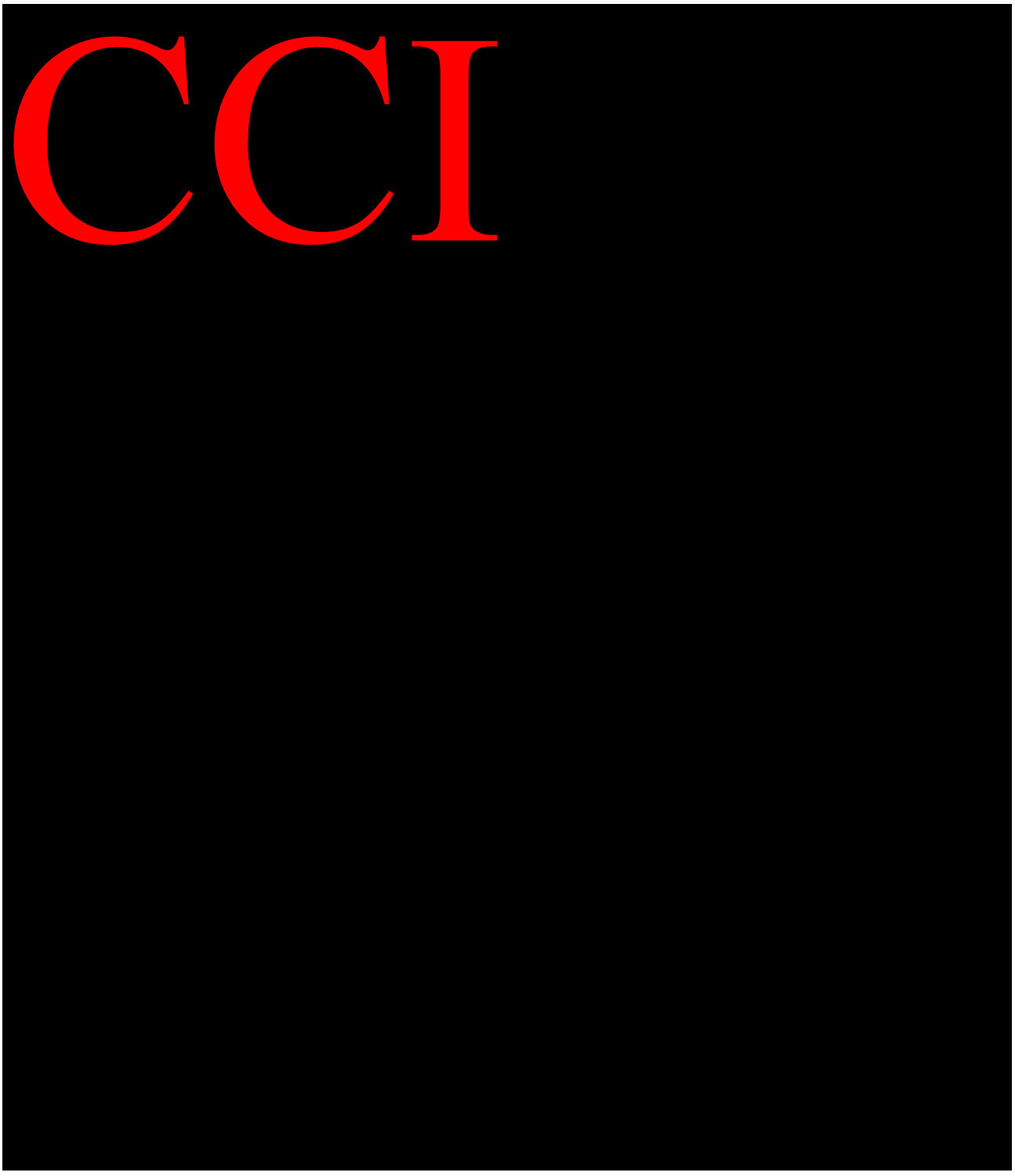
Tipping Point Analyses

A tipping point analysis will be performed on the primary and ALP normalization endpoints as binary variables. This analysis will explore the effect of missing data on the reliability of the efficacy results by determining the extent the missing data have to change for the results of the study to tip from statistically significant to not. The number of subjects in each treatment group with missing data that are categorized as non-responders will be incrementally decreased by 1 (and thus increasing the count of responders) to allow for construction of a heat map of p-values based on these incremental changes. This heat map will display p-values with color to indicate the magnitude of p-value departures from <0.05 significance by increases in the number of responders in each treatment group.

A similar method will be applied to NRS pruritus scores. To assess the robustness of the MMRM analysis results under MNAR (missing not at random) assumption, a delta-adjusting pattern-mixture approach for tipping point analysis, using the second variant based on the reference (Ratitch 2013) applied to each treatment group, will be conducted for the key secondary endpoint of change from baseline in weekly NRS pruritus. This method will be used to perform a series of analyses based on a range of different values of a shift parameter, δ, applied to each missing timepoint listed in Table 2 after MAR imputation of data to a monotone pattern for each treatment group. Pairs of δ values, belonging to those applied to placebo and seladelpar groups, respectively, can then be associated with various p-values based on the MMRM analyses.

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

MMRM analyses will be performed multiple times, using Rubin's rules (Rubin 1987) to construct p-values specific to each δ pair. A heat map, similar to that described for responder analyses, will display values which indicate the magnitude of p-value departures from



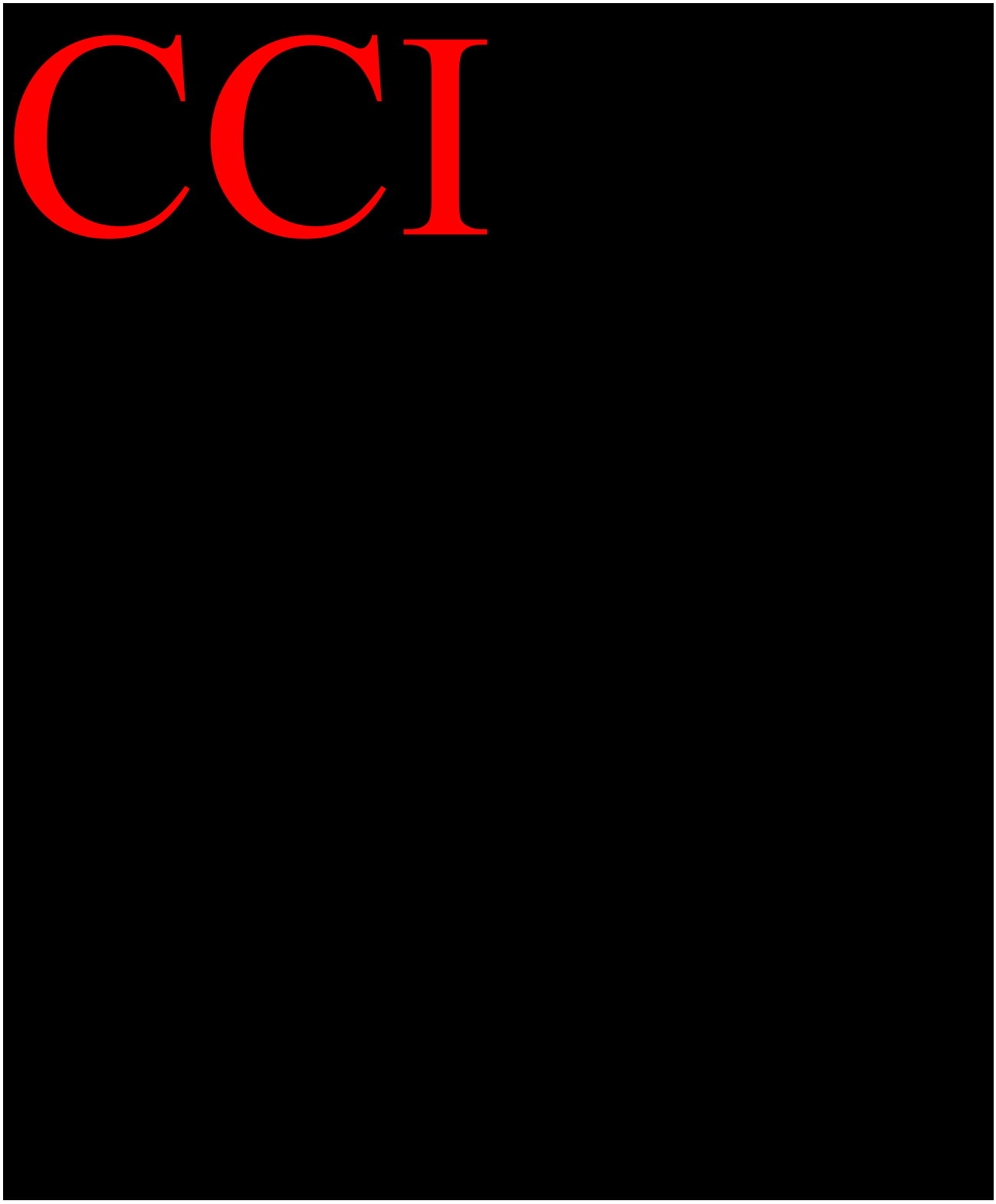
DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B



DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

The number of subjects who contributed PK samples at each timepoint will be summarized. Sample collection dates / time and concentration results will be listed. The PK analysis set will be used for tables and listings.

Further pooling of the concentration data from this study with data from other studies to facilitate development of a population PK model will be reported separately.

11. SAFETY

The purpose of this section is to describe the safety analyses for the study. Safety data will be summarized by actual treatment group and overall using the safety analysis set.

11.1. Adverse Events

An AE includes any condition resulting from concomitant illnesses, reactions to concomitant medications, or progression of disease states. Pregnancy should be documented as an AE and should be reported to the clinical monitor and to the Sponsor immediately upon learning of the event. Pregnancies will be followed through delivery or termination of the pregnancy.

Adverse events will be analyzed in terms of treatment-emergent adverse events (TEAEs); for the purposes of the SAP, the window for TEAEs will include any AE that starts after initiation of the study drug until up to 30 days after the last study medication administration. For the summaries of treatment-related AEs, a treatment-related TEAE is defined as an AE which was considered to be related (reported as “possible”, “probably”, or “definite” on the Adverse Events CRF page) to any study drug. AEs with a missing relationship will be presented in the summary table as a relationship category of “treatment-related”.

Adverse events will be coded by SOC and PT using MedDRA. The dictionary version used for reporting the study will be described in the relevant table and listing footnotes. The severity of AEs will be graded based on NCI CTCAE, Version 5.0, November 2017.

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

An overall summary of the number and percentage of subjects with any TEAE, serious TEAE, Grade 3 or above TEAE, treatment-related TEAE, treatment-related serious TEAE, treatment-related Grade 3 or above TEAE, TEAE leading to permanent withdrawal of study drug, treatment-related TEAE leading to permanent withdrawal of study drug, TEAE leading to study discontinuation, treatment-related TEAE leading to study discontinuation, TEAEs leading to dose interruptions, TEAEs leading to dose reductions, TEAEs leading to death, and treatment-related TEAE leading to death, will be provided by treatment group.

Tables summarizing the incidence of TEAEs in the Safety analysis set will be generated for each of the following categories, presented by treatment group:

- All TEAEs:
  - Presented by SOC and PT
  - Presented by PT
  - Occurring in ≥ 5% of subjects in any treatment group by PT
  - Leading to permanent withdrawal of study drug by SOC and PT
  - Leading to study discontinuation by SOC and PT
  - With fatal outcome by SOC and PT
- Grade 3 or higher TEAEs:
  - Presented by SOC, PT, and CTCAE grade
  - Presented by PT
- Serious TEAEs:
  - Presented by SOC and PT
  - Presented by PT
- Treatment-related TEAEs:
  - Presented by SOC and PT
  - Presented by PT
  - Grade 3 or higher, presented by SOC, PT, and CTCAE grade
  - Leading to permanent withdrawal of study drug by SOC and PT
  - Leading to study discontinuation by SOC and PT

Incidence tables will present TEAEs by SOC, PT, or both, sorted by decreasing frequency of the number (n) and percentage of subjects (%) in the seladelpar group. Counting of AEs will be by subject, and subjects will be counted only once within each SOC or PT. For tables categorized by severity, subjects with multiple events within a particular SOC or preferred term will be counted under the category of their most severe event with that SOC or preferred term. If the severity is missing for all occurrences of the AE, the subject will be counted only once in the 'Missing' category for severity. For the purpose of inclusion in TEAE tables, incomplete AE onset dates will be imputed based on methods described in Section 5.7.

Subgroup analyses (using the subgroups listed in Section 4.2) will be presented for the following categories of adverse events, presented by treatment group:

- Overall summary of all TEAEs
- All TEAEs by PT
- Treatment-related TEAEs by PT

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

The following TEAEs will be presented by preferred term, and by preferred term and CTCAE grade:

- Pruritus-Related TEAEs: Any TEAEs with preferred term containing "prur"
- Adverse events of interest will be presented for the following categories, by pre-defined search strategy:
    - TEAEs potentially reflecting liver-related toxicity (Hepatic disorders SMQ – Appendix 14.7)
    - TEAEs potentially reflecting muscle-related toxicity (Appendix 14.6)
    - TEAEs potentially reflecting renal-related toxicity (Acute renal failure SMQ – Appendix 14.8)
    - TEAEs potentially reflecting pancreatic-related toxicity (Appendix 14.9)
- Cardiovascular-Related TEAEs: Any TEAEs potentially reflecting cardiovascular-related toxicity (Cardiac arrhythmias SMQ, Cardiac failure SMQ, Cardiomyopathy SMQ, Ischaemic heart disease SMQ – Appendix 14.10)

Listings will be presented that include the verbatim term, preferred term, and SOC as well as full details of all AEs for all subjects in the Safety analysis set. Non-TEAEs will be flagged on this listing. A comprehensive listing of all AEs will be provided in a by-subject data listing. A listing of fatal outcomes will be provided for the safety analysis set.

In addition, listings will be provided for the following categories of TEAEs:

- Grade 3 or higher
- Serious
- Treatment-related and Serious
- Fatal Outcome
- Leading to
    - Dose interruptions/reductions
    - Permanent withdrawal of study drug
    - Study discontinuation
    - Fatal outcome
- Pruritus-related
- Potentially reflecting toxicity that is:
    - Liver-related
    - Muscle-related
    - Renal-related
    - Pancreatic-related
- Cardiovascular-related

11.1.1. TEAEs Associated with Safety Monitoring Criteria

As described in Section 10 of the protocol, subjects with specific lab abnormalities will be monitored closely and may interrupt study drug or discontinue study drug. A listing will be provided that includes subjects who meet laboratory safety monitoring criteria and for whom there is an associated TEAE reported.

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

11.2. Clinical Laboratory Evaluations

11.2.1. Hematology and Biochemistry

The following hematology parameters will be collected at Screening, Run-in, Day 1 and each post-baseline study visit: red blood cells (RBC, ie erythrocytes); hemoglobin; hematocrit; white blood cells (WBC; ie, leukocytes); WBC differential (absolute and percentage) including neutrophils, lymphocytes, basophils, eosinophils, and monocytes; platelets, prothrombin time, and INR. Platelets, prothrombin time, and INR will also be performed locally if deemed necessary by the Investigator.

The following biochemistry parameters will be collected at Screening, Run-in, Day 1 and each post-baseline study visit: ALP, AST, ALT, GGT, protein, albumin, total bilirubin, direct and indirect bilirubin, 5'-nucliotidase, aldolase, sodium, potassium, chloride, bicarbonate, blood urea nitrogen (BUN)/urea, serum creatinine, eGFR, CK, glucose, LDH, TG, total cholesterol, HDL-C, LDL-C, cystatin C, troponin I, lipase, and amylase.

Boxplots will display laboratory measures including amylase, albumin, ALT, AST, ALP, GGT, total bilirubin, creatinine, CK, lipase, eGFR. Line plots will display laboratory parameters over time for ALP, total bilirubin, direct bilirubin, indirect bilirubin, AST, ALT, GGT, 5'nucleotidase, platelets, INR, creatine kinase, creatinine, amylase, lipase, albumin, aldolase, and eGFR.

The eGFR will be calculated using the following MDRD equation:

$$\text{eGFR (mL/min/1.73 m2)} = 175 \times (Scr, std)^{-1.154} \times (Age)^{-0.203} \times (0.742 if female) \times (1.212 if African American)$$

where, $S_{cr,std}$ = serum creatinine in mg/dL measured with a standardized assay.

Scheduled biochemistry and hematology parameters will be graded according to NCI CTCAE version 5.0.

The following additional rules will be applied to CTCAE grading:

- Values that do not fall into the CTCAE grading range will be assigned a grade of 0, i.e Normal.
- Laboratory parameters that require clinical symptoms or interventions for grading will be graded without respect to this additional clinical information. Refer to Appendix 14.11 for details.
- For neutrophil, the CTCAE Grade 1 range falls within the central laboratory's normal reference range, Grade 1 may not exist. (see Appendix 14.11)
- For eGFR, the CTCAE grade 1 range was adjusted from <LLN – 60 mL/min/1.73m^2 to <90 – 60 mL/min/1.73m^2 due to the laboratory's reference range having a lower limit of normal of 60 mL/min/1.73m^2

All safety laboratory parameter data will be provided in subject data listings. Separate listings will be provided for the following laboratory assessments:

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

- Subjects with maximum post-baseline laboratory assessments of CTCAE ≥ Grade 3
- Subjects with maximum post-baseline laboratory assessments for selected laboratory parameters related to liver disease of CTCAE ≥ Grade 3
- Subjects who met safety monitoring criteria
- All abnormal laboratory hematology and chemistry values
- Local laboratory values

The following will be summarized in tables by treatment arm:

- Observed values and changes from baseline for hematology and biochemistry evaluations (except Troponin I) and model for end-stage liver disease (MELD) score for each visit, last observed value (LoV), and the maximum and minimum post-baseline values across all subjects and all time points.
- Shift from baseline to worst post-baseline value for hematology and biochemistry evaluations;
- Shift from baseline to minimum post-baseline value in total bilirubin, by selected ranges (< 0.6× ULN, ≥ 0.6 – ≤ 1× ULN, > 1× ULN)
- Shift from baseline to lowest and highest post-baseline value for bicarbonate, direct bilirubin, indirect bilirubin, aldolase, BUN, and cystatin C by category (low, normal, high)

11.2.2. Hematology and Biochemistry Parameters of Interest

Laboratories of interest listed in Table 5 below are intended to assess potential drug-induced liver injury (DILI), potential muscle injury, renal safety, and pancreatic safety. Subject incidence of these laboratory findings will be summarized.

Table 5: Laboratories of Interest

Laboratory Category	Laboratory Findings
Potential DILI	≥ 3×-, 5×-, 10×-, and 20× ULN elevations of ALT or AST
	Elevated TB to > 2× ULN
	Elevation of ALT or AST (> 3× ULN) accompanied by elevated TB (> 1.5× ULN, > 2× ULN)
	Elevation of ALT or AST > 3× ULN and (TB > 2× ULN or INR > 1.5)
	Elevation of ALT or AST > 3× ULN with the appearance of fatigue, nausea, vomiting, right upper quadrant pain or tenderness, fever, rash, and/or eosinophilia (> 5%)
	Baseline ALT and AST > ULN, and postbaseline ALT or AST > 2× baseline
	ALP ≥ 2× baseline
	ALP ≥ 2× baseline and TB ≥ 2× baseline
	ALP ≥ 3× baseline

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

Laboratory Category	Laboratory Findings
Potential muscle injury	CK (U/L) >3× ULN
	CK (U/L) >5× ULN
	CK (U/L) >10× ULN
Renal safety	Creatinine, increase (mg/dL) ≥ 1.5× baseline
	Creatinine, increase (mg/dL) ≥ 2.0× baseline
	Creatinine, increase (mg/dL) ≥ 3.0× baseline
	eGFR, decrease (ml/min/1.73m^2) ≥ 25% decrease
	eGFR, decrease (ml/min/1.73m^2) ≥ 50% decrease
	eGFR, decrease (ml/min/1.73m^2) ≥ 75% decrease
Pancreatic safety	Amylase (U/L) > 1.1× ULN
	Amylase (U/L) > 1.5× ULN
	Amylase (U/L) > 3.0× ULN
	Lipase (U/L) > 1.1× ULN
	Lipase (U/L) > 1.5× ULN
	Lipase (U/L) > 3.0× ULN

Abbreviations: ALT=alanine aminotransferase; AST=aspartate aminotransferase; CK=creatine kinase; DILI=drug-induced liver injury; eGFR=estimated glomerular filtration rate; INR=international normalized ratio; TB=total bilirubin; ULN=upper limit of normal.

11.2.3.4 Hepatocellular Drug-Induced Liver Injury Screening Analyses

Cases of possible drug-induced liver injury (DILI) will be identified in plots using maximum postbaseline total bilirubin and maximum postbaseline ALT and AST (ie, eDISH plots). Reference lines will be placed at 3× ULN for ALT and AST, and 2× ULN for total bilirubin. Interpretation of this screening plot will use a 4-quadrant approach. Listings with AST, ALT, and total bilirubin results will be provided for subjects falling into left upper quadrant, right lower quadrant, and right upper quadrant (possible Hy's Law quadrant).

A listing of potential Hy's law cases will be provided to identify cases that fall into the right upper quadrant in which postbaseline total bilirubin elevation to ≥ 2× ULN has occurred on or within 30 days after a postbaseline ALT or AST elevation to ≥ 3× ULN, without regard for ALP level.

11.2.3. Other Laboratory Assessments

HBsAg, HCV RNA, HIV, and local COVID-19 testing will be collected as noted in the schedule of assessments. These will not be presented in tables or listings. Serum pregnancy (β-HCG) and urine pregnancy tests will be listed.

11.3. Other Safety Evaluations

11.3.1. Vital Signs

Vital signs will be recorded at every study visit. Systolic and diastolic blood pressure, heart rate, respiratory rate, temperature, and weight will be collected. Body mass index (BMI) will be derived from weight and height.

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

Summary tables presenting observed values and change from baseline by treatment group will be presented for systolic blood pressure, diastolic blood pressure, temperature, heart rate, respiratory rate, weight, and BMI at baseline and each post-baseline visit. The last observed value (LoV), and maximum and minimum post-baseline values, along with change from baseline, will be presented. International System of Units will be applied in presenting the results.

Incidences of values of potential clinical concern for vital signs, as identified in Table 6 and Table 7, will be summarized using the most extreme change in each direction for each parameter.

Table 6: Vital Sign Values of Potential Clinical Concern

Vital Sign Parameter	Unit	Clinical Concern Range	
		Lower	Upper
Systolic Blood Pressure	mmHg	< 85	>160
Diastolic Blood Pressure	mmHg	< 45	> 100
Heart Rate	beats/min	< 40	> 110

Table 7: Vital Sign Changes from Baseline of Potential Clinical Concern

Vital Sign Parameter	Unit	Clinical Concern Range	
		Decrease	Increase
Systolic Blood Pressure	mmHg	≥ 40	≥ 40
Diastolic Blood Pressure	mmHg	≥ 20	≥ 20
Heart Rate	beats/min	≥ 30	≥ 20

A listing will be provided with all vital sign data by subject.

11.3.2. Electrocardiogram

A 12-lead electrocardiogram (ECG) was obtained in supine position after at least 5 minutes of rest at Screening, Day 1, Month 6, Month 12, and Follow-up visits, as well as at ET and unscheduled visits, if applicable. ECG interpretations (normal or abnormal) were collected, along with the specific abnormalities if the interpretation was abnormal.

Tables will be provided depicting the following, by treatment group:

- Descriptive statistics of ECG parameters (heart rate, PR, QRS, QT, and QTcF) at baseline and at each post-baseline timepoint, along with changes from baseline

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

- Number and percent of subjects with postbaseline QTcF values meeting criteria for potential clinical concern:
  - By range of observed QTcF value (≥ 450 msec, ≥ 480 msec, and ≥ 500 msec);
  - By categorical change in QTcF value from baseline (≥ 30 msec and ≥ 60 msec)

QTcF values that are missing on the eCRF will be derived as follows:

$$RR\ (sec) = \frac{60}{Heart\ Rate\ (beats/min)}$$

$$QTcF(msec) = \frac{QT(msec)}{(RR(\text{sec}))^{\frac{1}{3}}}$$

All recorded and derived ECG data will be presented by subject in a listing.

11.3.3. Physical Examinations

Physical examinations will be presented in subject data listings.

11.3.4. Abdominal Ultrasound

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

CCI

The protocol defines TEAE as any AE that newly appeared, increased in frequency, or worsened in severity after initiation of the study drug until up to 17 days after the last study medication administration, based on protocol v4, in which the safety follow-up visit occurs up to 14 days after the EOT visit. Previous versions of the protocol had a safety follow-up visit 1 month after the EOT visit. For the purposes of data summarization as outlined in this SAP, events up to 30 days after last dose of study medication will be summarized as TEAEs, capturing AEs from the safety follow-up visit for subjects under prior protocol versions.

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

13. REFERENCES

Elman S, Hynan LS, Gabriel V, and Mayo MJ. The 5-D itch scale: a new measure of pruritus. Br J Dermatol. 2010;162(3):587-93.

International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH E3 Guideline: Structure and content of clinical study reports questions & answers (R1). 2012. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E3/E3_QAs_R1_Step4.pdf. Accessed 13 Jan 2016.

International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline: Structure and content of clinical study reports E3. Step 4. 1995. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E3/E3_Guideline.pdf. Accessed 13 Jan 2016.

International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH E9: Statistical principles for clinical trials. 1998. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf . Accessed 15 June 2016.

International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use. E14 Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs. 2005. https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E14/E14_Guideline.pdf. Accessed 17 February 2017.

Hamilton S, Bernstein AB, Blakey G, Fagan V, Farrow T, Jordan D, Seiler W, Shannon A, Gertel A, Budapest Working Group: Developing the Clarity and Openness in Reporting: E3-based (CORE) Reference user manual for creation of clinical study reports in the era of clinical trials transparency. Research Integrity and Peer Review. (2016) 1:4.

PhUSE CSS Development of Standard Scripts for Analysis and Programming Working Group. Analyses and Displays Associated with Thorough QT/QTc Studies. 2016. http://www.phusewiki.org/docs/CSS%20White%20Papers%202016/230316%20CS_WhitePaper_TQTStudies_v1.0.pdf . Accessed 03 Jan 2017.

Liublinska, Viktoriia. 2013. Sensitivity Analyses in Empirical Studies Plagued with Missing Data. Doctoral dissertation, Harvard University.

Lammers, Willem J., et al. "Development and validation of a scoring system to predict outcomes of subjects with primary biliary cirrhosis receiving ursodeoxycholic acid therapy." Gastroenterology 149.7 (2015): 1804-1812.

Mantel, N., and Fleiss, J. L. (1980). “Minimum Expected Cell Size Requirements for the Mantel-Haenszel One-Degree-of-Freedom Chi-Square Test and a Related Rapid Procedure.” *American Journal of Epidemiology* 112:129–134.

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

Miettinen, O. S., and Nurminen, M. M. (1985). "Comparative Analysis of Two Rates." *Statistics in Medicine* 4:213–226.

Ratitch, B., Lipkovich, I., O'Kelly, M. "Combining Analysis Results from Multiply Imputed Categorical Data" Proceedings of PharmaSUG 2013 (Pharmaceutical Industry SAS Users Group), SP03, Baltimore.

Ratitch B and O'Kelly M, 2011, Implementation of Pattern-Mixture Models Using Standard SAS/STAT Procedures, PharmaSUG
https://www.pharmasug.org/proceedings/2011/SP/PharmaSUG-2011-SP04.pdf

Ratitch B, O'Kelly M, Tosiello R. "Missing Data in Clinical Trials: From Clinical Assumptions to Statistical Analysis Using Pattern Mixture Models." Pharm Stat. 2013;12:337-347.

Rubin DB (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons.

Santner, T. J., and Snell, M. K. (1980). "Small-Sample Confidence Intervals for $p_1 - p_2$ and p_1/p_2 in 2×2 Contingency Tables." *Journal of the American Statistical Association* 75:386–394.

Yan, X., Lee, S., Li N. (2009) "Missing data handling methods in medical device clinical trials. Journal of Biopharmaceutical Statistics" 19: 1085-1098.

Ye, J. (2016) General AUC calculated based on the trapezoidal rule
https://www.pharmasug.org/proceedings/china2016/DS/PharmaSUG-China-2016-DS02.pdf

Corpechot C, Carrat F, Poujol-Robert A, Gaouar F, Wendum D, Chazouillères O, Poupon R. Noninvasive elastography-based assessment of liver fibrosis progression and prognosis in primary biliary cirrhosis. Hepatology. 2012 Jul;56(1):198-208. doi: 10.1002/hep.25599. Epub 2012 Jun 5. PMID: 22271046.

Jones D, Carbone M, Invernizzi P, Little N, Nevens F, Swain MG, Wiesel P, Levy C. Impact of setanaxib on quality of life outcomes in primary biliary cholangitis in a phase 2 randomized controlled trial. Hepatology Communications. 2023 Mar;7(3).

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

14. APPENDICES

14.1. PBC Response Criteria

Criteria	Definition of Response
Barcelona	> 40 % decrease in ALP or ALP ≤1× ULN
Paris I	ALP < 3.0× ULN, AST < 2.0× ULN and total bilirubin ≤ 1 mg/dL
Paris II	ALP < 1.5× ULN, AST < 1.5× ULN, and total bilirubin ≤ 1 mg/dL
Toronto I	ALP ≤ 1.67× ULN
Toronto II	ALP ≤ 1.76× ULN
Rotterdam	Normalization of abnormal total bilirubin and albumin

14.2. UK-PBC Risk Score

The UK-PBC Risk Score calculator originally used information from the UK-PBC Research Cohort (Carbone et al., 2016) to estimate the risk (expressed in percentage) that a PBC subject established on treatment with UDCA will develop liver failure requiring liver transplantation within 5, 10 or 15 years from diagnosis. The UK-PBC risk score calculation is defined as follows:

UK-PBC risk score = 1 – baseline survivor function^exp (0.0287854*(ALP12-1.722136304) – 0.0422873*(((TA12/10^-1) – 8.675729006) + 1.4199*(ln(bilirubin12/10) + 2.709607778) – 1.960303*(albumin-1.17673001) – 0.4161954*(platelet-1.873564875)),

where ALP12, TA12 and bilirubin 12 refers to the ALP assessment, transaminases (refers to the ALT, where available, otherwise the AST) assessment and total bilirubin assessment, respectively, at the Baseline and corresponding post-baseline scheduled visits divided by the upper limits of their corresponding normal ranges; albumin and platelet represent their baseline assessment divided by their corresponding lower limits of normal ranges.

The baseline survivor function will take values 0.982 (at 5 years); 0.941 (at 10 years); and 0.893 (at 15 years).

14.3. GLOBE Risk Score

The GLOBE risk score (Lammers et al 2015) was developed to predict transplantation-free survival of UDCA treated subjects with PBC. The GLOBE risk score can be calculated as follows:

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

GLOBE score = (0.044378 * age + 0.93982 * LN(total bilirubin/ULN) + (0.335648 * LN(alkaline phosphatase/ULN)) – 2.266708 * albumin /LLN – 0.002581 * platelet count per 10^9/L) + 1.216865.

14.4. MELD Score

MELD(i) score = 10*[0.957 x LN(creatinine mg/dL) + 0. 378 x LN(bilirubin mg/dL) + 1.120 x LN (INR) + 0.643].

All laboratory values are rounded to 10th decimal place when calculating the score, and MELD(i) is rounded to the nearest whole number.

All laboratory values less than 1.0 will be set to 1.0 when calculating a candidate's MELD(i) score. Laboratory values for creatinine greater than 4.0mg/dL will be set to 4.0 mg/dL; sodium values less than 125 mmol/L will be set to 125 and sodium values greater than 137 mmol/L will be set to 137.

If MELD(i) is less than or equal to 11 then MELD = MELD(i).

If MELD(i) is greater than 11 then MELD = MELD(i) + (1.32 *(137 - (Na)) – (0.033*MELD(i)*(137 – Na))

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

14.6. TEAEs Potentially Reflecting Muscle-related Toxicity

Muscle related toxicity TEAEs is defined using broad Myalgia FMQ (FDA Medical Queries) v2.1:

PT	Final Classification
Chest tenderness	Narrow
Eosinophilia myalgia syndrome	Narrow
Fibromyalgia	Narrow
Fibromyalgia syndrome	Narrow
Fibrositis	Narrow
Musculoskeletal discomfort	Narrow
Myalgia	Narrow
Myalgia aggravated	Narrow
Myalgia intercostal	Narrow
Neuromuscular pain	Narrow
Polymyalgia	Narrow
Polymyalgia aggravated	Narrow
Polymyalgia rheumatica	Narrow
Buttock pain	Broad
Chest wall pain	Broad
Costal pain	Broad
Cramp-fasciculation syndrome	Broad
Intercostal pain	Broad
Lupus myositis	Broad
Muscle spasms	Broad
Muscle tightness	Broad
Musculoskeletal chest pain	Broad
Musculoskeletal pain	Broad
Musculoskeletal stiffness	Broad
Myofascial pain syndrome	Broad
Myositis	Broad
Rhabdomyolysis	Broad
Shoulder blade pain	Broad
Myofibrillar myopathy	Broad
Pregnenolone deficiency	Broad

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

14.7. TEAEs Potentially Reflecting Liver-related Toxicity

Liver related TEAEs is defined using broad Hepatic disorders SMQ (Standardised MedDRA Queries) v24.0, with exclusion of those sub-SMQs: Congenital, familial, neonatal and genetic disorders of the liver; Liver infections; and Pregnancy-related hepatic disorders.

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

14.8. TEAEs Potentially Reflecting Renal-related Toxicity

Renal related TEAEs is defined using broad Acute renal failure SMQ (Standardised MedDRA Queries) v24.0.

PT	Scope	Category
Acute kidney injury	Narrow	A
Acute phosphate nephropathy	Narrow	A
Anuria	Narrow	A
Azotaemia	Narrow	A
Continuous haemodiafiltration	Narrow	A
Dialysis	Narrow	A
Foetal renal impairment	Narrow	A
Haemodialysis	Narrow	A
Haemofiltration	Narrow	A
Neonatal anuria	Narrow	A
Nephropathy toxic	Narrow	A
Oliguria	Narrow	A
Peritoneal dialysis	Narrow	A
Prerenal failure	Narrow	A
Renal failure	Narrow	A
Renal failure neonatal	Narrow	A
Renal impairment	Narrow	A
Renal impairment neonatal	Narrow	A
Subacute kidney injury	Narrow	A
Albuminuria	Broad	A
Blood creatinine abnormal	Broad	A
Blood creatinine increased	Broad	A
Blood urea abnormal	Broad	A
Blood urea increased	Broad	A
Blood urea nitrogen/creatinine ratio increased	Broad	A
Creatinine renal clearance abnormal	Broad	A
Creatinine renal clearance decreased	Broad	A
Creatinine urine abnormal	Broad	A
Creatinine urine decreased	Broad	A
Crystal nephropathy	Broad	A
Fractional excretion of sodium	Broad	A
Glomerular filtration rate abnormal	Broad	A
Glomerular filtration rate decreased	Broad	A
Hypercreatininaemia	Broad	A
Hyponatriuria	Broad	A
Intradialytic parenteral nutrition	Broad	A
Kidney injury molecule-1	Broad	A
Nephritis	Broad	A

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

Neutrophil gelatinase-associated lipocalin increased	Broad	A
Oedema due to renal disease	Broad	A
Protein urine present	Broad	A
Proteinuria	Broad	A
Renal function test abnormal	Broad	A
Renal transplant	Broad	A
Renal tubular disorder	Broad	A
Renal tubular dysfunction	Broad	A
Renal tubular injury	Broad	A
Renal tubular necrosis	Broad	A
Tubulointerstitial nephritis	Broad	A
Urea renal clearance decreased	Broad	A
Urine output decreased	Broad	A

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

14.9. TEAEs Potentially Reflecting Pancreatic-related Toxicity

Pancreatic related toxicity TEAEs is defined using broad Pancreatitis FMQ v1.0:

PT	Final Classification
Alcoholic pancreatitis	Narrow
Alcoholic pancreopathy	Narrow
Autoimmune pancreatitis	Narrow
Cytomegalovirus pancreatitis	Narrow
Haemorrhagic necrotic pancreatitis	Narrow
Ischaemic pancreatitis	Narrow
Lupus pancreatitis	Narrow
Obstructive pancreatitis	Narrow
Oedematous pancreatitis	Narrow
Pancreas infection	Narrow
Pancreatic abscess	Narrow
Pancreatic haemorrhage	Narrow
Pancreatic necrosis	Narrow
Pancreatic phlegmon	Narrow
Pancreatic pseudocyst	Narrow
Pancreatic pseudocyst drainage	Narrow
Pancreatitis	Narrow
Pancreatitis acute	Narrow
Pancreatitis bacterial	Narrow
Pancreatitis chronic	Narrow
Pancreatitis due to biliary obstruction	Narrow
Pancreatitis fungal	Narrow
Pancreatitis haemorrhagic	Narrow
Pancreatitis helminthic	Narrow
Pancreatitis necrotising	Narrow
Pancreatitis relapsing	Narrow
Pancreatitis viral	Narrow
Pancreatorenal syndrome	Narrow
Traumatic pancreatitis	Narrow
Amylase abnormal	Broad
Amylase increased	Broad
Blood amylase abnormal	Broad
Blood amylase increased	Broad
Blood trypsin increased	Broad
Cullen's sign	Broad
Grey Turner's sign	Broad
Hereditary pancreatitis	Broad
Hyperamylasaemia	Broad
Hyperlipasaemia	Broad

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

Lipase abnormal	Broad
Lipase increased	Broad
Lipase urine increased	Broad
Pancreatic enzyme abnormality	Broad
Pancreatic enzymes abnormal	Broad
Pancreatic enzymes increased	Broad
Peripancreatic fluid collection	Broad
Ultrasound pancreas abnormal	Broad

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

14.10. TEAEs Potentially Reflecting Cardiovascular-related Toxicity

Cardiovascular related TEAEs is defined using broad Cardiac arrhythmias SMQ, Cardiac failure SMQ, Cardiomyopathy SMQ, Ischaemic heart disease SMQ v24.0.

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

14.11. Amended CTCAE for Applicable Laboratory Tests

Some laboratory tests include additional clinical criteria that cannot be determined programmatically. The table below specifies the grade with clinical criteria in *Italics*, per CTCAE version 5. The text in *Italics* will not be considered when assigning laboratory grades programmatically. Parameters not listed in this table will be graded without alteration of CTCAE rules.

Amylase (serum amylase increase) • Grade 1: >ULN - 1.5×ULN • Grade 2: >1.5×ULN - 2.0×ULN; *>2.0 – 5.0×ULN and asymptomatic* • Grade 3: >2.0 – 5.0×ULN *with signs or symptoms; >5.0×ULN and asymptomatic* • Grade 4: >5.0×ULN *and with signs or symptoms*
Bicarbonate (Blood bicarbonate decreased): • Grade 1: <LLN *and no intervention initiated*.
Prothrombin Time (PT)/International Normalized Ratio (INR) (INR increased): • Grade 1: >1.2 - 1.5; *>1 - 1.5 x baseline if on anticoagulation; monitoring only indicated* • Grade 2: >1.5 - 2.5; *>1.5 - 2.5 x baseline if on anticoagulation; dose adjustment indicated* • Grade 3: >2.5; *>2.5 x baseline if on anticoagulation; bleeding*
Lipase (Lipase increased) • Grade 1: >ULN – 1.5 × ULN • Grade 2: >1.5-2.0 × ULN, *>2.0-5.0 × ULN and asymptomatic* • Grade 3: >2.0-5.0 × ULN *with signs or symptoms; > 5 × ULN and asymptomatic* • Grade 4: >5.0 × ULN *with signs or symptoms*
eGFR (Chronic Kidney Disease): • Grade 1: <LLN - 60 ml/min/1.73 m^2 *or proteinuria 2+ present*; *urine protein/creatinine > 0.5* [LLN is defined as 90 ml/min/1.73 m^2 for this study, based on KDIGO 2012 Clinical Practice Guideline.] • Grade 2: 59 - 30 ml/min/1.73 m2 • Grade 3: 29 - 15 ml/min/1.73 m2 • Grade 4: <15 ml/min/1.73m2; *dialysis or renal transplant indicated*

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

Glucose (Hypoglycemia): • Grade 1: <LLN - 55 mg/dL; <LLN - 3.0 mmol/L • Grade 2: <55 - 40 mg/dL; <3.0 - 2.2 mmol/L • Grade 3: <40 - 30 mg/dL; <2.2 - 1.7 mmol/L • Grade 4: <30 mg/dL; <1.7 mmol/L; *life-threatening consequences; seizures*.
Potassium (Hypokalemia) • Grade 1: <LLN - 3.0 mmol/L • Grade 2: *Symptomatic with <LLN - 3.0 mmol/L; intervention indicated* • Grade 3: <3.0 - 2.5 mmol/L; *hospitalization indicated* • Grade 4: <2.5 mmol/L; *life-threatening consequences*
Potassium (Hyperkalemia) • Grade 1: >ULN – 5.5 mmol/L • Grade 2: >5.5 - 6.0 mmol/L; *intervention initiated* • Grade 3: >6.0 - 7.0 mmol/L; *hospitalization indicated* • Grade 4: >7.0 mmol/L; *life-threatening consequences*
Sodium (Hyponatremia) • Grade 1: <LLN - 130 mmol/L • Grade 2: 125-129 mmol/L *and asymptomatic* • Grade 3: *125-129 mmol/L symptomatic;* 120-124 mmol/L *regardless of symptoms* • Grade 4: <120 mmol/L; *life-threatening consequences*
Sodium (Hypernatremia) • Grade 1: >ULN-150 mmol/L • Grade 2: >150 - 155 mmol/L; *intervention initiated* • Grade 3: >155 - 160 mmol/L; *hospitalization indicated* • Grade 4: >160 mmol/L; *life-threatening consequences*
Albumin (Hypoalbuminemia) • Grade 1: <LLN - 3 g/dL; <LLN - 30 g/L • Grade 2: <3 - 2 g/dL; <30 - 20 g/L • Grade 3: <2 g/dL; <20 g/L • *Grade 4: Life-threatening consequences; urgent intervention indicated.*

DocuSign Envelope ID: 056E936E-990A-4E27-BEF1-D6872909798B

Magnesium (Hypomagnesemia) • Grade 1: <LLN - 1.2 mg/dL; <LLN - 0.5 mmol/L • Grade 2: <1.2 - 0.9 mg/dL; <0.5 - 0.4 mmol/L • Grade 3: <0.9 - 0.7 mg/dL; <0.4 - 0.3 mmol/L • Grade 4: <0.7 mg/dL; <0.3 mmol/L; *life-threatening consequences*
Magnesium (Hypermagnesemia) • Grade 1: > ULN – 3.0 mg/dL; ULN – 1.23 mmol/L • Grade 3: >3.0 – 8.0 mg/dL; > 1.23 -3.30 mmol/L • Grade 4: >8.0 mg/dL; >03.30 mmol/L; *life-threatening consequences.*
Triglycerides (Hypertriglyceridemia) • Grade 1: 150 mg/dL - 300 mg/dL; 1.71 mmol/L - 3.42 mmol/L • Grade 2: >300 mg/dL - 500 mg/dL; >3.42 mmol/L - 5.7 mmol/L • Grade 3: >500 mg/dL - 1000 mg/dL; >5.7 mmol/L - 11.4 mmol/L • Grade 4: >1000 mg/dL; >11.4 mmol/L; *life-threatening consequences*
Calcium (Hypocalcemia) • Grade 1: Corrected serum calcium of <LLN - 8.0 mg/dL; <LLN - 2.0 mmol/L; Ionized calcium <LLN - 1.0 mmol/L • Grade 2: Corrected serum calcium of <8.0 - 7.0 mg/dL; <2.0 - 1.75 mmol/L; Ionized calcium <1.0 - 0.9 mmol/L; *symptomatic* • Grade 3: Corrected serum calcium of <7.0 - 6.0 mg/dL; <1.75 - 1.5 mmol/L; Ionized calcium <0.9 - 0.8 mmol/L; *hospitalization indicated* • Grade 4: Corrected serum calcium of <6.0 mg/dL; <1.5 mmol/L; Ionized calcium <0.8 mmol/L*; life-threatening consequences*
Neutrophil count (Neutrophil count decreased) • Grade 1: LLN – 1500/mm3; < LLN - 1.5 x 10e9 /L [Depending on the central lab normal range (eg, Medpace LLN 1000/mm3), Grade 1 may not exist] • Grade 2: < 1500-1000/mm3; < 1.5 - 1.0 × 10e9 /L • Grade 3: < 1000-500/mm3; < 1.0 - 0.5 × 10e9 /L • Grade 4: < 500/mm3; < 0.5 × 10e9 /L