STATISTICAL ANALYSIS PLAN

Unique Protocol ID: CB8025-21629

Official Title: An 8-week, dose ranging, open label, randomized, Phase 2 study with a 44-week extension, to evaluate the safety and efficacy of MBX-8025 in subjects with Primary Biliary Cholangitis (PBC) and an inadequate response to or intolerance to ursodeoxycholic acid (UDCA)

NCT number: NCT02955602

Version Number: Version 4.1

Date Of SAP: 26-Nov-2019

Statistical Analysis Plan

NCT02955602

Sponsor:	CymaBay Therapeutics, Inc.
Protocol No:	CB8025-21629
Version No./Date	4.2/20 July 2017
Title	An 8-week, dose ranging, open label, randomized, Phase 2 study with a 44-week extension, to evaluate the safety and efficacy of MBX-8025 in subjects with Primary Biliary Cholangitis (PBC) and an inadequate response to or intolerance to ursodeoxycholic acid (UDCA)
CRF Version No./Date	4.1 19 Dec 2017
PRA Project Id:	CMY21629-21629X
SAP Version No./Date:	2.2 26 November, 2019 Final

Approvals

Sponsor	
Sponsor Name:	CymaBay Therapeutics, Inc.
Representative/Title:	[redacted]
Signature/Date:	[redacted]
PRA	
Project Manager/Title:	[redacted]
Signature/Date:	
Biostatistician/Title:	
Signature/Date:	

Table of Contents

1.0 Introduction

This statistical analysis plan (SAP) describes the statistical methods to be used during the reporting and analyses of data collected under CymaBay Therapeutics, Inc. Protocol CB8025-21629, titled "An 8-week, dose ranging, open label, randomized, Phase 2 study with a 44-week extension, to evaluate the safety and efficacy of MBX-8025 in subjects with Primary Biliary Cholangitis (PBC) and an inadequate response to or intolerance to ursodeoxycholic acid (UDCA)".

This SAP should be read in conjunction with the study protocol and case report form (CRF). This version of the plan has been developed using protocol versions 4.0 (United Kingdom version), 4.1 (German version), 4.2 (United States version), and 4.3 (Canadian version) dated 20 July 2017 and CRF version 4.1 dated 19 Dec 2017. Any further changes to the protocols or CRF may necessitate updates to the SAP. Changes following the approval of the SAP will be tracked in the SAP Change Log and a final version of the SAP will be issued for sponsor approval prior to database lock.

2.0 Study Objectives

2.1 Primary Objective

The primary objective of this study is to evaluate the safety and efficacy of MBX-8025 2 mg, 5 mg, and 10 mg over 8 weeks of treatment.

2.2 Secondary Objectives

The secondary objectives for this study are:

- To evaluate the safety and efficacy of MBX-8025 2 mg, 5 mg, and 10 mg over 12 and 26 weeks of treatment
- To evaluate the safety and efficacy of MBX-8025 2 mg, 5 mg, and 10 mg over 52 weeks of treatment
- To evaluate the pharmacokinetics of MBX-8025

2.3 Exploratory Objective

The exploratory objectives of this study are to evaluate the effect of MBX-8025 on bile acids, additional markers of inflammation and renal function.

3.0 Study Design

This will be an international, multicenter study in males and females with PBC who have inadequate response or intolerance to UDCA.

The study will utilize an open-label, partially randomized, uncontrolled, 8-week dose ranging parallel groups (2 mg, 5 mg, and 10 mg) design with an open-label extension for a total of up to 52-weeks of treatment.

After signing an informed consent form (ICF), subjects will enter a screening period (up to 2 weeks). Subjects in the 5 mg and 10 mg treatment groups will enter the 8-week initial treatment period study in parallel. Subjects in the 2 mg group will enter the 8-week initial treatment period in chronological order.

Randomization will occur only for 5 mg and 10 mg treatment groups; subjects in the 2 mg treatment group will be registered for the study. Randomization/registration will occur ideally at least 7 days prior to first dose of MBX-8025 (study drug). A subject will be considered formally entered in the trial at randomization/registration. Study drug kits will be shipped from the drug deposit facility to the site on a subject-by-subject basis. Subject-specific study drug supplies (MBX-8025) will be taken orally once a day, for a period of 8 weeks. To confirm study drug compliance, MBX-8025 trough level will be measured after 2 and 8 weeks of treatment.

After completion of the 8-week initial treatment period, subjects will enter the open-label extension for a total of up to 52 weeks of treatment. Subjects will initially enter the extension on their assigned dose. Based

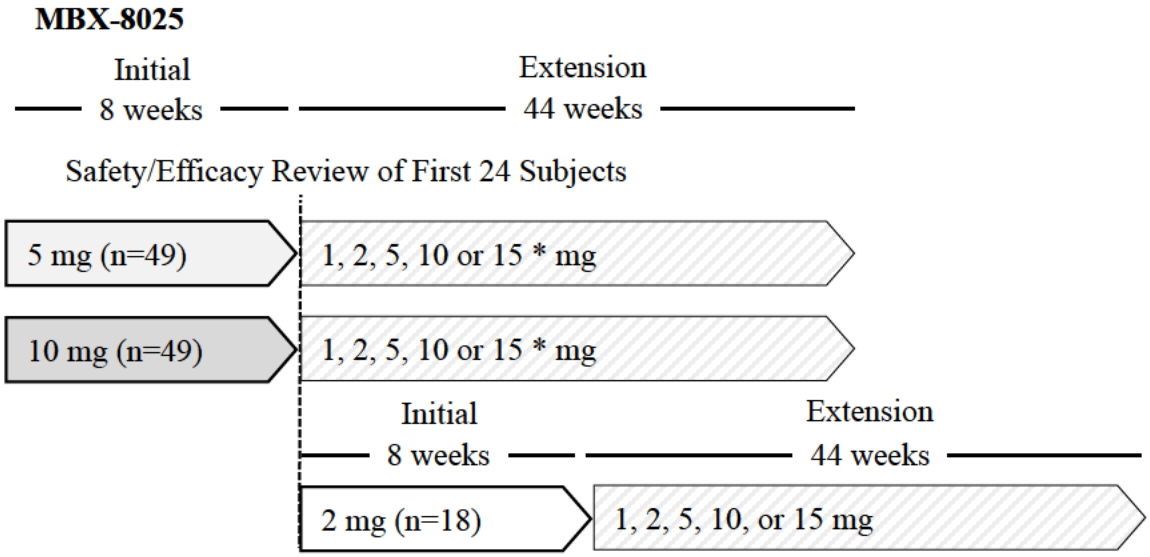
on this individual subject review and the overall study safety and efficacy review, the dose might be up- or down-titrated for the extension part of the study. During the extension, the subject's dose might be re-adjusted for safety or efficacy reasons. For individual subjects in the 2 mg, 5 mg, and 10 mg treatment groups, an ongoing safety and efficacy review will determine the dose to be used (1 mg, 2 mg, 5 mg, 10 mg, or 15 mg. Subject's safety and efficacy review will be performed by the Investigator in collaboration with Medical Monitor. During the extension period, subject-specific study drug supplies (MBX-8025) will be taken orally once a day, for a total maximum duration of up to 52 weeks.

Four interim data reviews were conducted while the study is ongoing. See section 7.1 for more details.

Final SAP1 v1.1 dated 14 June 2017 was used to perform these interim analyses.

After the end of treatment, subjects will enter a four-week follow-up period.

At selected centers subjects will participate in 24-hour PK testing. 24-hour PK testing will be performed at the time of initiation of study treatment and after 2 weeks or 12 weeks of treatment. Subjects participating in 24-hour PK testing will have PK blood and urine samples, and bile acids samples collected multiple times within 24 hours.



* US-only: Dose up-titration to 15 mg can be performed only within 26-week treatment period. Beyond Week 26 (Visit 11), the highest dose will be 10 mg.

Table 1: Schedule of Assessments

Visit	V1	V2	V3	V4	V5	V6	V7	Post Dose Adjust[12]	V8	V9	V10	V11	V12	V13	V14/ EOT	V15/ EOS	ET	UNS
Week	W -2 to Day 1	Day 1	W1	W2	W4	W6	W8		W12	W16	W20	W26	W32	W39	W52	W56		
Visit/Study Periods	Screening	Baseline	Initial Treatment					Extension								Follow-up		
Informed Consent	X																	
Eligibility Evaluation	X																	
Randomization[1]	X																	
Medical History[2]	X																	
Physical Exam[3]	X	X[4]	X[4]	X[4]	X[4]	X[4]	X		X[4]	X[4]	X[4]	X	X[4]	X[4]	X	X	X	X[4]
Vital Signs and Weight	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Height	X																	
ECG[5]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Hematology[6]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Biochemistry[6,7]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Exploratory Biochemistry[6,8]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Bile Acids/Sterols		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Serum Pregnancy Test	X	X			X		X		X	X	X	X	X	X	X	X	X	
Back-up Serum Sample		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
PK Trough Level				X			X											
24-hour Blood Samples (selected centers only)[9]		X							X									
24-hour Urine Samples (selected centers only)[10]		X							X									
Urinalysis (spot)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
5D-itch Scale		X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	
Pruritus VAS		X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	
PBC-40 QoL		X			X		X		X	X	X	X	X	X	X	X	X	
AE/TEAE		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Prior and Concomitant Medications	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
MBX-8025 Intake On-site		X	X	X	X	X	X	X	X	X	X	X	X	X	X			
MBX-8025 Compliance			X	X	X	X	X	X	X	X	X	X	X	X	X		X	X
MBX-8025 Accountability			X	X	X	X	X	X	X	X	X	X	X	X	X		X	X
MBX-8025 Dispensing		X			X		X		X	X	X	X	X	X				
Safety/Efficacy Review[11]							X											

1. Randomization will occur only for 5 mg and 10 mg treatment groups. Subjects in the 2 mg treatment group will be registered for the study. Randomization/registration will occur ideally at least 7 days prior to Visit 2 (Baseline). Study drug kits will be shipped from the drug depot facility to the site on a subject-by-subject basis.
2. Including PBC history
3. Liver-related symptoms will be evaluated as part of physical exam. UNS visit might be scheduled, if needed per PI judgment.
4. Symptoms-directed (brief) physical examination
5. ECG will be performed approximately 60-90 minutes after dosing (Visit 2 through Visit 11)
6. Blood will be collected after at least an 8-hour overnight fast and prior to dosing. If the subject forgets to fast, the site will continue to draw labs
7. If at Visit 1 (Screening) an unexpected abnormal CK level is observed, re-test the subject to confirm eligibility
8. At Visit 1 (Screening) only AMA is to be performed
9. The 24-hour blood sample (selected centers) will be collected at the following time points: within 30 minutes prior to dosing, and 30 min, 1, 2, 4, 6 and 24 hours after the dosing
10. The 24-hour urine sample (selected centers) will be collected over the following intervals: 0-6 hours and 6-24 hours.
11. Safety and efficacy data review will be performed after a subject completes the initial 8-week treatment period. The review will be done by the Investigator in collaboration with Medical Monitor. Based on this review, the dose might be up- or down-titrated for the extension period.
12. Applicable only for subjects who had dose adjustment at any time during extension. The visit will occur in two weeks after initiation of the extension dose. If dose adjustment occurs approximately two weeks prior to the next scheduled visit, post-dose adjustment visit and the next planned visit can be combined.

3.1 Sample Size Considerations

Approximately 116 subjects will participate in the study:

- 2 mg treatment group: up to 18 subjects
- 5 mg treatment group: approximately 49 subjects
- 10 mg treatment group: approximately 49 subjects

PK (24-hour): up to 30 subjects

- 2 mg treatment group: up to 6 subjects
- 5 mg, and 10 mg treatment group: up to 12 subjects per group.

It is assumed that the 5 mg and 10 mg treatment groups will have at least 10% difference in the mean AP percent change with a 15 % standard deviation. Based on this assumption, and on the use of two-sided, two-sample t-test at the alpha =0.05 level of significance, a study sample size of 49 subjects per group will have a 90% power to detect a 10% mean difference between 5 mg and 10 mg treatment groups. There is no formal sample size justification for 2 mg treatment group.

3.2 Randomization

Subjects in the 5 mg and 10 mg treatment groups will be centrally randomized in a 1:1 ratio to receive MBX-8025 5 mg or MBX-8025 10 mg. Subjects in the 2 mg treatment group will be registered to the study.

4.0 Study Variables and Covariates

4.1 Efficacy

4.1.1 Primary Variable

- Percent change from baseline to end of 8-week treatment in AP level.

4.1.2 Secondary Variables

- Change from baseline to 12 weeks and 52 weeks of treatment in AP level.
- Percent change from baseline to 12 weeks and 52 weeks of treatment in AP level.
- Composite responder endpoint of AP and total bilirubin:
  - AP < 1.67 × upper limit of normal (ULN) and
  - Total Bilirubin within normal limit and
  - $\geq$ 15% decrease in AP
- Change and percent change from baseline to 12 weeks and 52 weeks of treatment in the following parameters:
  - Aspartate aminotransferase (AST), alanine aminotransferase (ALT), gamma-glutamyl transferase (GGT), 5'nucleotidase, bilirubin (total, conjugated, unconjugated), bone-specific AP, Triglycerides (TG), total cholesterol (TC), high density lipoprotein Cholesterol (HDL-C), and low density lipoprotein cholesterol (LDL-C), PBC response criteria (Barcelona, Paris I and II, Toronto I and II)
  - UK-PBC risk score
  - 5D-itch scale

- Pruritus Visual Analog Score (VAS)
- PBC-40 QoL
- Model for End-Stage Liver Disease (MELD) scores
- GLOBE score

4.1.3 Exploratory Variables

- Exploratory biochemistry: Anti-Mitochondrial Antibodies (AMA), C-terminal telopeptide (CTX), cystatin C, IgM, fibrinogen, fibroblast growth factor 19 (FGF19), haptoglobin, high-sensitivity C-Reactive Protein (hs-CRP), homocysteine, microRNA-122, and N-terminal propeptide of type 1 procollagen (P1NP);
- Bile acids: Primary Bile Acids and Salts: Cholic acid (CA), chenodeoxycholic acid (CDCA), glycocholic acid (GCA), taurocholic acid (TCA), glycochenodeoxycholic acid (GCDCA), taurochenodeoxycholic acid (TCDCA); Secondary Bile Acids and Salts: Deoxycholic acid (DCA), lithocholic acid (LCA), glycodeoxycholic acid (GDCA), glycolithocholic acid (GLCA), taurodeoxycholic acid (TDCA), taurolithocholic acid (TLCA); Other Bile Acids: ursodeoxycholic acid (UDCA), glycoursodeoxycholic acid (GUDCA), tauroursodeoxycholic acid (TUDCA);
- Sterols: desmosterol (DESM), lanosterol (LANO), lathosterol (LATH), 7- dehydrocholesterol (7-DHC), cholestanol (CSTN), coprostanol (COPR), squalene (SQLN); β-sitosterol (β-SITO), campesterol (CAMP), stigmasterol (STIG), 7α-hydroxycholesterol (7 α-HC);
- Bile acid precursor: C4 (7α-hydroxy-4-cholesten-3-one);

4.2 Pharmacokinetics (PK)

4.2.1 General Methods

24-hour plasma and urine concentration determination of MBX-8025 and its metabolites (M1, M2 and M3) will be performed on subjects participating in 24-hour PK testing. 24-hour blood samples will be collected at Visit 2 (Day 1) and Visit 4 (Week 2) or 8 (Week 12). Blood samples will be collected within 30 minutes prior to dosing, and 30 min, 1, 2, 4, 6 and 24 hours after dosing.

24-hour urine sample will be collected over two intervals 0-6 hours and 6-24 hours.

PK trough level: Plasma concentration determination of MBX-8025 and its metabolites (M1, M2 and M3) will be collected from non-PK subjects only pre-dose at Visits 4 (Week 2) and Visit 7 (Week 8).

Sample collection, processing and handling details are provided in the Laboratory Manual.

4.2.2 Bioanalysis

Samples will be analyzed at a laboratory approved by the sponsor and stored at a facility designated by the sponsor.

Concentrations of MBX-8025 and its metabolites (M1, M2, and M3) will be assayed using validated analytical methods.

4.2.3 Handling of Missing Data for PK Analyses

No data exclusions or imputations are planned. Exclusions or imputations may be made in the case of implausible or outlier data. Such cases will be adjudicated by a trained pharmacokineticist and reported with justification. In cases where imputations or exclusions may affect the conclusions or interpretation of the results significantly, an assessment of such impact may be conducted, at the discretion of the pharmacokineticist, by comparing the results with and without such actions. All imputations and

exclusions need to be approved by the sponsor upon discussing the justification, and imputations and exclusions will be documented in the report along with the justification.

4.2.4 Plasma Concentrations

Plasma concentrations of MBX-8025 and its metabolites (M1, M2 and M3) that below the limit of quantification (BLQ) before t_{max} will be treated as zero and after t_{max} will be treated as missing. BLQ values embedded between two measurable values will be treated as missing.

Linear and semi-log plots of the mean plasma concentration by scheduled sampling time will be provided by treatment. These plots will show time in hours. The plots will match the summary table results.

For calculation of area under the plasma concentration curve (AUC), BQL values on the leading edge of the profile (i.e., before T_{max}) are set equal to zero in the dataset loaded into WinNonlin (WNL) for pharmacokinetic analysis. BQL values at the end of a profile are also set to missing.

4.2.5 Pharmacokinetic Data Analysis Methods

Pharmacokinetic parameter estimates for MBX-8025 and its metabolites (M1, M2 and M3), in plasma will be calculated by PRA by standard noncompartmental methods of analysis using WinNonlin Phoenix version 8.1 or higher (Certara, Princeton, NJ, USA).

Descriptive statistics (n, mean, SD, %CV, median, min, max, geometric mean, and geometric CV) of PK parameters will be presented for each study dose.

The following plasma pharmacokinetic parameters will be calculated for MBX-8025 and its metabolites (M1, M2 and M3) on both Day 1 (Visit 2) and Week 2 (Visit 4) Week 12 (Visit 8).

Subjects' MBX-8025 and metabolite concentrations with time of plasma sample collection relative to dosing (collection time minus dose time (hours)) and PK parameters will be listed.

Parameter	Description	SAS Programming Notes
C_{max}	Maximum plasma concentration. Observed peak analyte concentration obtained directly from the experimental data without interpolation, expressed in concentration units	Cmax from WNL
T_{max}	Time to maximum plasma concentration. First observed time to reach peak analyte concentration obtained directly from the experimental data without interpolation, expressed in time units.	Tmax from WNL If needed, additional T_{max} will be added by SAS programming
AUCs	Area under the plasma concentration-time curve. Calculated using linear linear up log down method, expressed in units of concentration x time.	
$AUC_{0\text{-last}}$	Area under the concentration-time curve (time 0 to time of last quantifiable concentration, typically 24 hours for this study).	AUClast from WNL
$AUC_{0\text{-tau}}$	Area under the curve for one dosing interval (tau), where tau = 24 hours. If the last time point(s) is (are) missing data, then the 24 time point is extrapolated.	AUC_TAU from WNL

Parameter	Description	SAS Programming Notes
λ_z	Terminal phase rate constant calculated by linear regression of the terminal log-linear portion of the concentration versus time curve.	Lambda_z from WNL The following conditions need to be met to keep the estimate: Linear regression of at least three points, AUC_%Extrap_obs <=20%, Rsq > .80, and λ_z interval > $t_{1/2}$ are required to retain λ_z. λ_z and all dependent parameters will be reported, but will not be included in descriptive statistics, if the λ_z interval < 2x $t_{1/2}$
$t_{1/2}$	Terminal phase half-life expressed in time units.	HL_Lambda_z from WNL If AUC_%Extrap_obs >20%, Rsq ≤ .80 or λ_z interval < $t_{1/2}$, then parameter is deleted $t_{1/2}$ will not be included in summary statistics if λ_z interval < 2x $t_{1/2}$.
C_{trough}	Predose concentration prior to each daily dose during Week 2 (Visit 4) and Week 8 (Visit 7)	Calculated in SAS
Week 2 RA_{Cmax}	Accumulation ratio C_{max}, calculated as C_{max} Week 2/C_{max} Day 1	Calculated in SAS
Week 12 RA_{Cmax}	Accumulation ratio C_{max}, calculated as C_{max} Week 12/C_{max} Day 1	Calculated in SAS
Week 2 RA_{AUC}	Accumulation ratio AUC, calculated as $AUC_{tau, Week\ 2}/AUC_{tau, Day\ 1}$ where tau = 24 hours	Calculated in SAS
Week 12 RA_{AUC}	Accumulation ratio AUC, calculated as $AUC_{tau, Week\ 12}/AUC_{tau, Day\ 1}$ where tau = 24 hours	Calculated in SAS

24-hour urine samples will be collected over two intervals, 0-6 hours and 6-24 hours. The following urine PK parameters will be calculated for MBX-8025 and its metabolites (M1, M2 and M3) at each interval.

Parameter	Description	SAS Programming Notes
Ae_{t-x}	Amount excreted in urine over an x-hour period.	Calculated in SAS

Parameter	Description	SAS Programming Notes
	Calculated as concentration of pooled hour urine sample 0-6, 6-24 and 0-24 x urine volume for that collection interval Ae_{t-x} = (concentration of pooled urine collected from time = t to time = x) * volume of pooled urine collected from time = t to time = x). $Ae_{0-24} = Ae_{0-6} + Ae_{6-24}$	

4.3 Safety

4.3.1 Primary Variables

- AEs and TEAEs: The severity of AEs will be graded by investigators using NCI CTCAE version 4.03
- ECGs
- Laboratory tests (biochemistry, hematology)
- Urinalysis (spot, 24 hour)

4.3.2 Secondary Variables

- Vital signs
- Physical examination
- Use of concomitant medications

4.4 Predetermined Covariates and Prognostic Factors

The baseline AP assessment is considered as covariate in the analysis of covariance (ANCOVA) model for the primary analysis. In addition, for the secondary efficacy variable analyses, the baseline assessment is also treated as covariate in the corresponding ANCOVA models.

5.0 Definitions

5.1 General

Assigned Treatment Group/Initial Dose Group

The assigned treatment group is the treatment group to which subjects were either enrolled to (for the 2 mg treatment group) or randomized to (for the 5 mg and 10 mg treatment groups). For the purposes of the data presentations, this will be referred to as Initial Treatment group.

5 mg Cohort Group

The Cohort Treatment group represents only the subjects who were originally assigned to the 5 mg dose. These subjects will be summarized based upon whether the subject titrated up to the next dose level or not. The Cohort Treatment Group is defined as follows:

- 5 mg: This group will consist of all subjects randomized to the 5 mg treatment group who never uptitrated during the study.

- 5/10 mg titration: This group will consist of all subjects randomized to the 5 mg treatment group who uptitrated to 10 mg at any time during the study.

Age at PBC diagnosis

Age at PBC diagnosis is calculated as: (Onset date of PBC in medical history CRF page - the birth date+1)/365.25. See imputation of partial missing dates in Section 9.5.2 Methods for Handling Dropouts and Missing Data.

Baseline

Baseline is defined as the mean between all assessments at and between Visit 1 (Week -2 to Day 1) and Visit 2 (Day 1). In cases where some endpoints may not capture multiple observations prior to Day 1 the single collected observation will be used.

Change from Baseline

Change from baseline is defined as (value at post-baseline visit – baseline value [as defined above]).

Concomitant and Prior Medication

Prior medications are defined as medications with a stop date prior to first dose of MBX-8025. Concomitant medications are defined as any medications ongoing at the start of MBX-8025 or with a start date on or after the first dose date.

Duration of PBC

Duration of PBC in years is defined as (date of informed consent - onset date of PBC in medical history CRF page +1)/365.25. See imputation of partial missing dates in Section 9.5.2 Methods for Handling Dropouts and Missing Data. If imputation is used and the result is the same as the date of first dose, then add 6 months to the duration of PBC calculation.

End of Treatment

Study End of Treatment (EOT) is defined as the last on treatment visit, for most patients this is Visit 14 (Week 52). Subject EOT is defined as Date of Last Dose on CRF EOT form and is denoted EOT on the CRF.

Last observation carried forward (LOCF)

The imputation of data will only be used for the analysis of AP and the Composite Endpoint. Imputation where missing post-baseline data less than two days after EOT will be imputed by carrying forward the last non-missing on treatment post-baseline value (LOCF).

Week 8 Endpoint

The Week 8 Endpoint is defined as the Visit 7 (Week 8) response. If a subject discontinues from the study prior to Week 8 but after Week 6, the missing value for AP and Composite Endpoint will be imputed using LOCF.

Week 12 Endpoint

The Week 12 Endpoint is defined as the Visit 8 (Week 12) response. If a subject discontinues from the study prior to Week 12 but after Week 8, the missing value for AP and Composite Endpoint will be imputed using LOCF. Otherwise, the Week 12 Endpoint is missing.

Week 52 Endpoint

Study Week 52 Endpoint is defined as Visit 14 (Week 52). If a subject discontinues from the study prior to Week 52 but after Week 26, the values for AP and the Composite Endpoint will be imputed using LOCF. For all other endpoints no imputation will be performed.

Percent change from Baseline

The percent change from baseline is calculated as (value at post-baseline visit – baseline value [as defined above]) X 100 / (baseline value [as defined above]).

If the baseline value is 0 and the post-baseline value is also 0, then the percent improvement from baseline is set to 0. If the baseline value is 0 and the post-baseline value is non-zero, then the percent improvement from baseline is set to "missing" (or '.').

Study Day 1

The first day of MBX-8025 administered. For subjects who are enrolled (randomized or registered) but not dosed, study day 1 is defined as the date of enrollment.

Study Day

Study Day is defined as the number of days from Study Day 1.

- Before Study Day 1: Study Day = (Date of Interest – Date of Study Day 1)
- On or After Study Day 1: Study Day = (Date of Interest – Date of Study Day 1) + 1

Therefore, the day prior to Study Day 1 is -1.

Study Completer

Subjects will be considered to have completed the study if they complete Visit 14 (Week 52).

Study Enrollment (Randomization or Registration)

Study enrollment (randomization for 5 and 10 mg cohorts, registration for 2 cohort) is defined as when subject receives a treatment allocation via the IXRS system.

UDCA daily dose

UDCA daily dose (mg/day) = Visit UDCA dose (mg) x number of times per day.

UDCA dose by weight (mg/kg/day) at each visit = (Visit UDCA dose (mg) x number of times per day) / visit weight (kg).

UDCA compliance will be estimated based on the number of missed doses reported by the subject. UDCA compliance = (Total duration of study drug exposure in days – number of days UDCA doses missed)/ Total duration of study drug exposure in days.

Visit Dose

For presentation of study assessments, Visit Dose is defined as the MBX-8025 dose received at the previous visit, the dose related to the assessments.

Visit Windows

In general, CRF visits will be used for this study. Unscheduled visits will also be denoted as *last_visit.1/2/3.* Visit 14 is EOT and Visit 15 is Follow-up even if these (EOT and Follow-up) occur prior to the planned visit windows.

Unscheduled visits include only lab assessments and AEs. Unscheduled visit lab assessments will be included in subject listings but not tables. Unscheduled visit AEs will be included in both tables and listings.

AP Isozymes

Three AP isozymes (Liver, Bone and Intestinal) which make up total AP measurements are provided in terms of percentage of total by Quest laboratory. Absolute conversions will be calculated in order to perform summaries on absolute measurements of these isozymes of AP response. The calculations will be performed for the following clinical laboratory parameters:

- Bone-specific AP

- Liver AP
- Intestinal AP

The conversion is calculated as (proportion reported) X (Total AP). The Total AP used will be the result reported by Quest laboratories and the resulting units will be U/L.

5.2 Efficacy

5-D Itch

The 5-D itch scale was developed as a five-dimension questionnaire designed to assess itching in clinical trials. The five dimensions are degree, duration, direction, disability and distribution. Detailed information on this scale can be found at Appendix D in the protocol.

Single-item domain scores (duration, degree and direction) are equal to the value the response choice (range 1–5).

The disability domain comprises four items that assess the impact of itching on daily activities: sleep, leisure/social activities, housework/errands and work/school. The score for the disability domain is the highest score on any of the four items.

For the distribution domain, the number of affected body parts is tallied (potential sum 0–16) and the sum is sorted into five scoring bins: sum of 0–2 = score of 1, sum of 3–5 = score of 2, sum of 6–10 = score of 3, sum of 11–13 = score of 4, and sum of 14–16 = score of 5.

The scores of each of the five domains are summed together to obtain a total 5-D score. 5-D scores can potentially range between 5 (no pruritus) and 25 (most severe pruritus).

5-D Itch modified

A modified score using the sum of 4 domains, will exclude the single item domain Direction. The scores of the remaining 4 domains will be summed to obtain a 5-D modified total score ranging from 4 (no pruritus) to 20 (most severe pruritus).

This modification is due to the wording of the Direction item. Subjects with no pruritus often record “unchanged,” which has a score of 4 instead of “completely resolved,” a score of 1.

Pruritus VAS

The Pruritus VAS will be used as a secondary efficacy measure to evaluate pruritus in subjects with PBC. VAS has values (unit: mm) ranging from 0 to 100, where 0 means no itching and 100 represents worst possible itching. The scoring procedure can be found in Appendix E of the protocol.

PBC Response Criteria

The following biochemistry criteria will be used to measure the PBC subject response, where ULN refers to the upper limit of the normal range. Missing AP or total bilirubin at Initial and Extension Endpoint visits will be imputed by LOCF.

1. Paris I: AP ≤ 3x ULN and AST ≤ 2x ULN and Total Bilirubin ≤ 1 mg/dL;
2. Paris II: AP ≤ 1.5x ULN and AST ≤ 1.5x ULN and Total Bilirubin ≤ 1 mg/dL
3. Toronto I: AP ≤ 1.67x ULN
4. Toronto II:
   - AP ≤ 1.76x ULN
5. Mayo II

- AP ≤1.67x ULN and Total Bilirubin ≤ ULN

6. Barcelona:
   - Normalization of AP or a Decrease of AP ≥ 40%
7. Rotterdam:
   - Early: normal Total Bilirubin and normal Albumin
   - Moderately advanced: Either abnormal Albumin or abnormal Total Bilirubin
   - Advanced: Both abnormal Albumin and abnormal Total Bilirubin
   -

UK-PBC Risk Score

The UK-PBC Risk Score calculator originally used information from the UK-PBC Research Cohort 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. Similar UK-PBC risk score calculation is adapted for this trial with the formula as below:

UK-PBC risk score = 1- 0.982 ^EXP(0.0287854*(AP12 x ULN-1.722136304) - 0.0422873*(((TA12 x ULN/10)^-1) - 8.675729006) + 1.4199 * (LN(BIL12 x ULN/10)+2.709607778)-1.960303*(Albumin x LLN-1.17673001)-0.4161954*(Platelet x LLN-1.873564875))

AP12, TA12 and BIL12 refers to the AP, transaminases (refers to the ALT, where available, otherwise the AST), and total bilirubin assessments, respectively, at specific visits. (Note: where AP12, TA12 and BIL12 are specified, the value at that visit will be used.) Albumin and platelet represent their baseline assessment. Missing AP12, TA12 or BIL12 at Initial and Extension Endpoint visits will not be imputed.

Baseline survivor function will take values of 0.982, 0.941, and 0.893 for 5 years, 10 years and 15 years respectively.

PBC-40 QoL

PBC-40 is a disease-specific health related quality of life tool developed to specifically measure the psychometric profile of PBC subjects. It covers six domains relevant to PBC including cognitive, social, emotional function, fatigue, itch, and other symptoms. See Appendix F of the study protocol for details.

Within each domain, items are scored from 1 to 5 and the individual item scores are summed to give a total domain score. The direction of scoring of some items is reversed for calculation of domain scores so that in all cases, high scores represent high impact and low scores low impact of PBC on quality of life.

MELD and MELD-Na 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))

MELD-Na = MELD(i) – Na – [0.025 * MELD(i) * (140 – Na)] + 140

GLOBE Score

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.

5.3 Safety

Creatinine Clearance

$CrCl(mL/min/1.73\ m^2) = [\ U_{Cr}\ (mg/dL) \times U_{vol}\ (mL)] / [S_{Cr}\ (mg/dL) \times time\ (min)\ ] \times [\ (1.73) / BSA\ ]$

Where U=urine, S=serum, and

DuBois & DuBois Body Surface Area (BSA) = $0.20247 \times height\ (m)^{0.725} \times weight\ (kg)^{0.425}$

6.0 Analysis Sets

Each subject can be titrated to a higher or lower dose, after Week 8 and may have more than one dose change. All analyses will be performed using the treatment groups as defined in Section 5.

6.1 Safety Population

The safety population is defined as any enrolled (randomized or registered) subject who receives at least one dose of medication. All safety analyses will be completed using the safety population by the initial treatment group and 5 mg cohort groups as defined in Section 5.

6.2 Intention-to-Treat Population (ITT)

The ITT population is defined as any enrolled subject who receives at least one dose of medication and has at least one post-baseline AP evaluation. The ITT population will be used for analyzing the primary endpoint and selected secondary endpoints according to subjects' initial treatment group as defined in Section 5.

6.3 Modified intention-to-Treat Population (mITT)

The mITT population is defined as all subjects in the ITT population who have confirmed PBC. The mITT population is the primary analysis population for all efficacy analyses. The efficacy analyses will be performed using the initial dose group and 5 mg cohort groups.

6.4 Per Protocol Population (PP)

The PP population is defined as all subjects in the mITT Population whose study drug compliance and UDCA compliance are between >=80% and <=120%.. Supportive efficacy analyses will be completed using the PP population using the same treatment groups as the mITT analyses.

6.5 Pharmacokinetic Population (PK)

All subjects who receive at least one dose of MBX-8025 and have at least one measurable drug concentration will be included in PK analyses. Analyses will be completed using the study dose received. PK visits occur at Day 1 and Week 2 or Week 12 Treatment study period.

7.0 Interim and Final Analyses

Four interim data reviews were conducted while the study is ongoing as follows. No aggregate or individual subject level data was unblinded in presentation of analysis results at any time.

7.1 Interim Analyses

7.1.1 24 Subjects at 8 Weeks

After the 24th subject in the 5 mg and 10 mg treatment groups completed the 8-week initial treatment period, a safety and efficacy data review was performed to determine whether the 20 mg treatment group should be initiated and the dose(s) to be used in the extension (2 mg, 5 mg, 10 mg, or 20 mg, see Section 3.0 Study Design for titration details).

7.1.2 24 Subjects at 12 Weeks

After the 24th subject in the 5 mg and 10 mg treatment groups completed the 12-week endpoint period, a safety and efficacy data review was performed. For both interim analyses, only the first 24 subjects who qualify for the analysis were included in the efficacy results. This subject set was determined by the 24th subject to qualify for the first interim analysis.

7.1.3 45 Subjects at 26 Weeks

After the 45th enrolled subject had their Week 26 visit or early terminates from the study, a third interim analysis was conducted. Only the first 45 enrolled subjects were included in the analyses.

7.1.4 Subjects at 52 Weeks

After the 40th enrolled subjects had their Week 52 visit, a fourth interim analysis was conducted. Analyses were conducted for the following subject groups: 12 week completers (all subjects who had completed at least 12 weeks of treatment); 26 week completers (all subjects who have completed at least 26 weeks of treatment); and 52 week completers (all subjects who had completed at least 52 weeks of treatment);

8.0 Data Review

8.1 Data Handling and Transfer

Data will be entered and exported as SAS® version 9.4 datasets. Converted datasets will be created using SAS® and following standard Clinical Data Interchange Standards Consortium Standard Data Tabulation Model (CDISC SDTM, version 1.3, Implementation Guide version v3.1.3) conventions. Analysis datasets will be created using SAS® and following CDISC Analysis Data Model (ADaM, version 2.1, Implementation Guide 1.0) standards.

Medical history and AEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) version 22.1 at the time of the analysis to assign a system organ class (SOC) and preferred term (PT) to each event. Prior and concomitant medications will be coded using the WHO Drug 2019MAR03DDE B3 of the World Health Organization Collaborating Centre (WHOCC) at the time of the analysis.

Additional details can be found in the PRA Data Management Plan for this study.

8.2 Data Screening

Review of a pre-freeze TFL dry run allows for further data screening prior to data base freeze for the primary analysis. The activities planned for pre-freeze phase can be found in the Data Management Plan. The pre-freeze TFL will be discussed with the sponsor in a data review meeting to identify any final data issues and seek corrections prior to database freeze. Database lock will occur at the end of the study, after all follow-up visits are complete. The PRA statistician and the sponsor must approve database lock.

9.0 Statistical Methods

All statistical analyses will be performed using SAS® (Version 9.4 or higher).

The final analysis will be performed at the end of the study (end of trial) after all subjects have completed the Week 56/Early Termination assessments and the follow-up assessment or have entered the long-term safety study.

Unless otherwise specified, descriptive data summaries will be tabulated by treatment for all endpoints. Categorical data will be summarized using number of subjects (n), frequency and percentages of subjects falling into each category, with the denominator for percentages being the number of subjects in the study population for each treatment group, unless otherwise noted. Percentages will be rounded to one decimal place except for 100%, which will have no decimal place.

All continuous variables will be summarized using mean, standard deviation (SD), standard error (SE), 95% confidence interval of mean, median, minimum, maximum, and number of subjects with observations. The mean and its confidence interval, median, SD, and SE will be presented to one decimal place greater than the original data, and the minimum and maximum will have the same number of decimal places as the original data.

9.1 Subject Disposition

The following information will be summarized for subject disposition and accountability:

- Subject disposition (including number of subjects who were screened, randomized, treated with MBX-8025, completed treatment, discontinued treatment with reason of discontinuation, completed study, and discontinued study with reason of discontinuation). Number of subjects who have dosage adjustments due to adverse event (AE), defined as subjects who have checked 'Temporarily interrupted, restarted – lower dose' for 'Action taken with the study treatment' in AE CRF page, will be provided as well. Summary will be presented overall, the Week 12 Treatment Period, the Week 52 treatment period and overall.
- Summaries of analysis sets with reason for exclusion from the per-protocol analysis set for all enrolled (randomized and registered) subjects.

Disposition summaries will be presented by initial treatment and the Cohort treatment group.

9.2 Protocol Deviations

Protocol deviations (PD) data will be entered into the Clinical Trials Management System (CTMS). The study team will conduct ongoing reviews of the PD data from CTMS and the resulting set of subjects to be included in the PP population throughout the study, adjusting the PD criteria as appropriate. The subjects to be included in the PP population will be finalized prior to database lock.

Based on the PD data entered into CTMS, the important protocol deviations (IPD), which are defined in the PD guidance, thought to potentially impact the statistical analyses, will be listed and tabulated for the ITT population using incidence and percentages by deviation type and treatment group as defined in Section 5. This also applies to the specific IPDs that will exclude subjects from the per protocol analysis set.

9.3 Treatments

9.3.1 Extent of Study Drug Exposure

MBX-8025 will be supplied as 1 mg, 5 mg, and 10 mg capsules. MBX-8025 will be administered orally, once daily in doses of 2 mg, 5 mg, and 10 mg as described in the protocol. During the extension periods, dose may be up- or down-titrated based on individual subject review and the overall study safety and efficacy review as described in Section 3.0 Study Design.

Summary statistics, defined below, will be provided for total number of doses taken, total duration (days) of MBX-8025 exposure, treatment compliance and the number of subjects having dose interruption/discontinuations due to AEs by treatment group as defined in Section 5. The analysis will be performed using the safety population, for the first 12 weeks, and for the entire study period (up to 52 weeks).

Total duration of study drug exposure in days = date of last dose – date of first dose + 1.

Total number of capsules planned = (date of last dose – date of first dose +1) * the number of capsules per day. In the case that a subject takes multiple capsules per day, this will be included in the above calculation, e.g., two 5 mg tablets to achieve a daily dose of 10 mg.

Total number of capsules received = sum of (number of capsules dispensed – number of capsules returned) for each kit record in DA2 CRF page across all kits. Kits not returned will have their capsules excluded from this calculation.

Treatment compliance (%) = 100* (total number of capsules taken/ total number of capsules planned).

A subject listing of MBX-8025 accountability and a listing of exposure in the study will be provided.

A listing of subjects who titrated up or down (or both) at any time on study will be provided.

9.3.2 Prior and Concomitant Medications/Procedures

Prior and concomitant medications will be coded by World Health Organization Collaborating Centers (WHOCC) criteria and will be summarized by Anatomical Therapeutic Chemical (ATC) system class, generic name and treatment group as defined in Section 5. The analysis will be performed using the Safety population.

The number and percentage of subjects using each medication will be displayed by treatment. Subjects taking more than one medication in the same generic name or ATC class will be counted once.

A list of prior and concomitant medications and procedures will be presented. A list of UDCA medication at each visit will also be presented.

9.4 Demographic and Baseline Characteristics

Demographic and baseline characteristics including age, race, gender, ethnicity, height, weight and body mass index (BMI) will be summarized using descriptive statistics for continuous variables and frequency distributions for discrete variables for safety, and mITT populations by initial treatment group and 5 mg cohort groups as defined in Section 5.Age (in year) = (screening date – birth date +1)/365.25

BMI (kg/m^2) = weight (kg)/height (m)^2

Medical history at screening will be summarized by SOC and PT and tabulated by treatment group and total for safety analysis population.

PBC history will be summarized using the safety/ITT and mITT populations.

9.5 Efficacy Analyses

The primary efficacy analyses will be based on the mITT population using the initial treatment group defined in Section 5. The primary endpoint and secondary endpoints will also be analyzed on the ITT population according to subjects' initial treatment group. All statistical testing will be performed using two-sided tests at the alpha=0.05 level of significance.

Descriptive summary will be provided on the mITT population using the initial treatment group and 5 mg cohort groups

Selected analyses will be repeated using PP population where indicated in the text below.

9.5.1 Primary Variable

Primary Analysis

ANCOVA on percent change from baseline

The primary endpoint is the percent change in AP serum level (U/L) from the baseline to the Week 8 Endpoint (the 8 week visit (Visit 7), or end of treatment (EOT) if before 8 weeks). Primary analysis will be performed using mITT with missing assessments imputed by LOCF as defined in Section 5.1. The analyses

will assess the change from baseline in each treatment group and will assess the hypothesis that there are no differences in the percent change in AP serum level from baseline to 8 weeks between MBX-8025 treatment groups. Paired comparisons between the doses will be performed.

An ANCOVA model with treatment group as the main effect and the baseline AP assessment as a covariate will be used for the comparison of primary endpoint analysis. Sample SAS code for PROC MIXED is displayed below:

```
PROC mixed data=datain;
  CLASS treatment;
  MODEL percent change= treatment baseline / clparm;
   Lsmeans treatment/pdiff CL;
   Estimate 'MBX-8025 10 vs. 2' treatment -1 0 1;
   Estimate 'MBX-8025 10 vs. 5' treatment 0 -1 1;
   Estimate 'MBX-8025 5 vs.2' treatment  -1 1 0;
```

Note: Depending on treatment assignment coding, the contrast in the estimate statement may need to be revised.

From the model, the least square means (LS Means), SE, 95% confidence interval of LS Means in the percent change in AP serum level from baseline to the visits of interest for each treatment group, the differences of LS Means (SE) between each MBX-8025 dose, and the corresponding p-values will be presented. The visits of interest are: Weeks 8, 12 and 52. At Week 8 and 12 the comparisons will be made for the 3 initial dose groups, at Week 52 the comparison will be limited to the comparison of the initial dose groups of 5 and 10 mg. Note: that actual treatment through Week 12 will be the initial treatment at baseline since titration will not have occurred prior to that point.

A descriptive summary of observed AP values at each visit including baseline and change from baseline as well as the percent change from baseline at each visit will be provided by treatment group.

The above ANCOVA model and descriptive analysis will be repeated using the ITT population as supportive analyses.

The within subject difference will be presented and a paired t-test will be used to show the subject differences between the baseline value and the Week 12 and Week 52 visit.

Sensitivity Analyses

ANCOVA Model assumptions including normality of the standardized residuals and homogeneity of variance between initial treatment groups will be assessed for the Week 8 endpoint. If any model assumptions are violated to the degree that model performance is in question, the Wilcoxon Mann-Whitney test (presenting Wilcoxon p-values) will be used to do pair-wise comparisons as the primary efficacy analysis with the sample SAS code listed below. Paired comparisons between MBX-8025 doses will not be adjusted for multiple comparisons.

```
proc NPAR1WAY data=datain wilcoxon;
        title "Nonparametric test to compare percent change in AP between treatment groups";
        class treatment;
        var percent_change;
        exact wilcoxon;
  run;
```

To assess the robustness of the primary analysis results, the same analysis models using the mITT population will be repeated using the PP population.

To explore the impact of UDCA intolerant subjects the primary analysis will be repeated excluding UDCA intolerant subjects.

Repeated measures of percent change from baseline

AP percent change from baseline will also be analyzed based on a repeated measures analysis of covariance, where observed data from post-baseline visit assessments (no missing data imputation will be used) during the from baseline through Week 52 are included. Main effects will be treatment group and visit, an interaction of treatment and visit and baseline will be the covariate.

The mixed model repeated measures analysis and will be implemented using PROC MIXED procedure in SAS, specifying the variance-covariance structure with the type which assumes the least underlying data structure that is able to converge (ie unstructured [UN], auto-regressive 1 [AR(1)], and compound symmetry [CS]). Example code for PROC MIXED is displayed below:

```
Proc mixed data=datain;
    Class treatment visit subject;
    Model percent_change= treatment visit treatment*visit baseline;
    Repeated visit/subject=subject type=un/ar(1)/cs;
    Lsmeans treatment*visit/pdiff cl;
Run;
```

UDCA Intolerant Subjects

The effect of treatment on UDCA intolerant subjects may be explored by creating individual response profiles.

9.5.2 Methods for Handling Dropouts and Missing Data

For the primary analysis based on mITT, missing values will be imputed using the LOCF method except for the repeated measure analysis. As sensitivity analyses, primary efficacy endpoints will also be analyzed for the PP analysis sets based on Observed Values in a repeated measures model as described in Section 9.5.1.

Imputation for Partial or Missing Dates

If dates are missing or incomplete for an AE (including deaths) or concomitant medication, the following algorithm will be used for imputation:

Table 2. Imputation Rules for Partial or Missing Start Dates

Start Date		Stop Date						
		Complete: yyyymmdd		Partial: yyyymm		Partial: yyyy		Missing
		<1st dose	≥1st dose	<1st dose yyyymm	≥1st dose yyyymm	<1st dose yyyy	≥1st dose yyyy	
Partial: yyyymm	= 1st dose yyyymm	2	1	n/a	1	n/a	1	1
	≠ 1st dose yyyymm		2	2	2	2	2	2

Partial: yyyy	= 1st dose yyyy	3	1	3	1	n/a	1	1
	≠ 1st dose yyyy		3		3	3	3	3
Missing		4	1	4	1	4	1	1

1 = Impute as the date of first dose
2 = Impute as the first of the month
3 = Impute as January 1 of the year
4 = Impute as January 1 of the stop year
Note: If the start date imputation leads to a start date that is after the stop date, then there is a data error and do not impute the start date.

Imputation rules for partial or missing stop dates:

1. Initial imputation
    a. For partial stop date “mmyyyy”, impute the last of the month.
    b. For partial stop date “yyyy”, impute December 31 of the year.
    c. For completely missing stop date, do not impute.
2. If the stop date imputation leads to a stop date that is after the death date, then impute the stop date as the death date.
3. If the stop date imputation leads to a stop date that is before the start date, then there is a data error and do not impute the stop date.

Imputation rules for partial or missing death dates:

1. If death year and month are available but day is missing:
    d. If “mmyyyy” for last contact date = “mmyyyy” for death date, set death date to the day after the last contact date.
    e. If “mmyyyy” for last contact date < “mmyyyy” for death date, set death date to the first day of the death month.
    f. If “mmyyyy” for last contact date > “mmyyyy” for death date, data error and do not impute.
2. If both month and day are missing for death date or a death date is totally missing, set death date to the day after the last contact date.

The imputed dates will be used to assess whether AEs should be considered as treatment emergent and if medications should be included in the safety summaries as prior or concomitant, however the original, partial dates will be included in data listings.

9.5.3 Multiplicity

Paired comparisons between MBX-8025 doses will be performed at the alpha=0.05 level of significance with no adjustment for multiple comparisons.

9.5.4 Pooling of Sites

All sites will be pooled for the analyses and site effect will not be evaluated.

9.5.5 Secondary Variables

All secondary analyses will be carried out using two-sided tests at the alpha=0.05 level of significance for the mITT population by the initial treatment groups. Select secondary endpoints (composite endpoint, total bilirubin, AST, ALT, GGT) will also be analyzed on the ITT population by the initial treatment groups. Analyses will be restricted to Week 12 and 52, in a similar manner to that above for the Primary Endpoint.

9.5.5.1 Composite Endpoint of AP and Total Bilirubin

At each post-baseline visit, a subject is defined as a responder if their AP assessment and total bilirubin satisfy the composite endpoint definition, that is, AP<1.67 x upper limit of normal and a ≥15% decrease from baseline and total bilirubin within the normal limit. Two-sided p-values from the Fisher's exact tests comparing MBX-8025 doses will be reported each study visit. The number (%) of responders and non-responders will be displayed by visit and treatment group, Statistical comparisons will be restricted to Weeks 12 and 52.

This analysis will be repeated using the PP population.

This analysis will be repeated excluding UDCA intolerant subjects.

9.5.5.2 Biochemistry Measurements

The percent changes from baseline in the following secondary biochemistry parameters as shown in Section 4.1.2. Secondary Variables will be analyzed separately by a similar ANCOVA model as that used for the primary efficacy variable. The difference in LS Means of percent change from baseline between MBX-8025 levels as well as their p-values and 95% confidence intervals will be presented, if available.

9.5.5.3 Published PBC Response Criteria

For each published PBC response criteria defined in Section 5.2, subjects who satisfy the criteria will be defined as responders. Otherwise, subjects will be considered non-responders if all the elements have a non-missing assessment at that visit.

The number of non-responders of each published criteria will be descriptively summarized at baseline and Weeks 12 and 52. Fisher's exact tests will test the difference between the responder rates between MBX-8025 treatment groups separately for each criterion.

A listing of subject's responses to each published criterion will be provided.

9.5.5.4 5-D Itch

At Week 12 and Week 52, the change and percent change from baseline in total 5-D score and modified total 5-D score between MBX-8025 treatment groups will be analyzed by ANCOVA models similar to the description for the primary variable, with percent change from baseline in total 5-D score and modified total as response variables, treatment group as factor and baseline total modified total 5-D itch scores as covariate.

The total and modified total 5-D itch scores change from baseline and percent change from baseline will also be descriptively summarized at each visit.

The individual domain score and total score will be displayed for each visit in a listing.

9.5.5.5 Pruritus VAS

The actual value and change from baseline at each visit in Pruritus VAS will be descriptively presented by Initial treatment group as defined in Section 5. Similar analyses to those described in Section 9.5.5.4 for 5-D Itch will be conducted for Pruritus VAS change from baseline to Week 12 and Week 52 visits. A listing for Pruritus VAS measured at each visit will be provided.

Additionally, three scatterplots to show the correlations between total 5-D itch score verse Pruritus VAS score will be provided at baseline, Week 12 and Week 52.

9.5.5.6 PBC-40 QoL

The change and percent change from baseline for individual domain score at Week 12 and Week 52 will be analyzed by the similar model as described in Section 9.5.5.5 for 5-D Itch. The actual individual domain scores as well as the change and percent change from baseline will also be summarized by visit and treatment groups.

The individual domain score and total score will be displayed for each visit in a listing.

9.5.5.7 UK-PBC Risk Score

The UK-PBC risk score for 5, 10 and 15 years will be calculated per Section 5.2 and a descriptive summary for each dose will be presented with LOCF imputed values for missing laboratory parameters assessment at Week 12 and Week 52.

9.5.5.8 MELD Score

The MELD scores will be calculated per Section 5.2 and a descriptive summary for each treatment group will be presented.

9.5.5.9 GLOBE Score

The GLOBE scores will be calculated per Section 5.2 and a descriptive summary for each treatment group will be presented.

9.6 Safety Analyses

All safety analyses will be performed on the Safety population based on subject's initial treatment group and 5 mg cohort groups as defined in Section 5 and in total across treatment groups. Safety listings will also display the initial dose group as well as the last dose level taken prior to the measurement/event.

9.6.1 Adverse Events

An AE includes any condition that: 1) was not present prior to study treatment, but appeared or reappeared following initiation of study treatment, or 2) was present prior to study treatment, but worsened during study treatment. This would include any condition resulting from concomitant illnesses, reactions to concomitant medications, or progression of disease states. Pregnancy should be documented as an AE. AE severity will be graded from 1 to 5 according to the CTCAE version 4.0 criteria (V4.03, June 14. 2010).

A treatment-emergent AE is defined as an AE with onset date at or after the first study medication administration and up to 30 days after the last study medication administration. Treatment-related AEs are AEs with relationship to the study treatment characterized as 'Possible' or 'Related' in the CRF.

All reported AEs will be coded to the appropriate SOC and PT according to the version 22.1 of MedDRA. A listing of all AEs will be provided.

Subject incidence of the following AEs will be tabulated by treatment groups and by SOC and PT:

- treatment-emergent AEs
- treatment-related AEs
- treatment-emergent AEs leading to discontinuation of treatment
- treatment-emergent AEs leading to discontinuation from study

Subject incidence of the following AEs will be tabulated by preferred term in descending order of frequency:

- treatment-emergent AEs occurring in ≥5% of subjects in any treatment group
- grade 3 or higher treatment-emergent AEs

In addition, subject incidence of treatment-emergent AEs and treatment-related AEs will be tabulated by treatment and by SOC, PT, and maximum severity

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.

The following listings will be also provided: 1) subject listing of AEs; 2) subjects with AEs leading to discontinuation of treatment; 3) Subjects with AEs leading to discontinuation from study; 4) Subjects with AE leading to dose adjustment; 5) Subjects experiencing pruritus, and 6) Subjects who met monitoring criteria listed in Section 10.5.1 of the protocol:

Liver Safety Monitoring

- Elevation of ALT/AST

  Normal ALT/AST at baseline and

  - ALT/AST >5× ULN or
  - Total bilirubin is ≥ 1.5 × ULN or ≥ 2 mg/dL
- Elevated ALT/AST at baseline
  - ALT/AST 3 × baseline and
  - PT/INR > 1.5 × ULN
- Elevation of total bilirubin (>2 × ULN or > 1.5 × baseline), regardless of ALT or AST levels, and
  - eosinophilia > 5%

Muscle Safety Monitoring

I. CK > 5 × ULN

II. CK > 2.5 × ULN and ≤ 5 × ULN

Serum Creatinine Monitoring

I. Serum creatinine > 2.0 × ULN

II. Serum Creatinine > 1.5 × ULN and <2.0 × ULN

Pancreatic Safety Monitoring

I. Amylase > 3 × ULN and/or lipase > 3 × ULN

9.6.2 Deaths and Serious Adverse Events

Subject incidence of the following will be tabulated by SOC and PT:

- serious treatment-emergent AEs
- serious treatment-emergent treatment-related AEs

A subject listing of SAEs and AEs with outcome of fatal will be also provided.

9.6.3 Laboratory Data

Blood samples will be collected after at least an 8-hour overnight fast and prior to dosing. If the subject forgets to fast, the site will continue to draw labs.

Routine hematology and biochemistry and urinalysis (spot) will be assessed at all visits including ET (early termination) and UNS (unscheduled) if applicable. Urinalysis (24-hr) will be collected at Visit 2 (Day 1) and Visit 4 (Week 2) or Visit 8 (Week 12)., over two intervals 0-6 hours and 6-24 hours.

Laboratory results which are BQL will be given a value of the lower level of quantification. PK results which are BQL will be assigned values as described in Section 4.2.4.

Routine Hematology:
Erythrocyte Count (RBC), Hemoglobin, Hematocrit, Leukocyte Count (WBC), WBC Differential (absolute and percentage), Platelets, PT/INR.

Routine biochemistry:
AP, AST, ALT, GGT, 5' nucleotidase, Protein, Albumin, Total Bilirubin, Conjugated Bilirubin, Troponin I, Unconjugated Bilirubin, Bone-specific AP (absolute units), Liver AP(absolute units), Intestinal AP(absolute units), Aldolase, Sodium, Potassium, Chloride, Bicarbonate, BUN/Urea, Creatinine, eGFR, CK (if above upper limit of normal, CK-MM, CK-MB, and CK-BB will be measured), venous blood glucose, LDH, TG, TC, HDL-C, and LDL-C, Amylase, Creatine, Lipase.

Shift tables of the most extreme post-baseline value (CTCAE grade 1-5, or low/normal/high) relative to baseline will be presented by treatment for parameters that have both non-missing baseline and post-baseline values.

Additional biochemistry measurements will be analyzed as described in Section 9.7 Exploratory Analyses and are listed in Section 4.1.3 Exploratory Variables.

Urinalysis (spot):
Sample will be analyzed via dip stick. Blood, Glucose, Ketones, Leukocyte Esterase, Nitrites, pH, and Protein will be measured. If positive, microscopic examination of sediments (Urinary White Blood Cells, Urinary Red Blood Cells) will be performed. Urine myoglobin will be measured if muscle injury is suspected.

Urinalysis (24-hr):
Will be analyzed for urine creatinine, and creatinine clearance (24-hr) will be calculated.

Conventional units and ranges will be used for reporting. Out of range of 'High' and 'Low' will be evaluated. Shift tables of worst post-baseline value relative to baseline will be presented by treatment for parameters that have both non-missing baseline and post-baseline values.

Laboratory variables that are included in Section 4.1 Efficacy, will be analyzed as described in section 9.5 Efficacy Analyses. Descriptive summary will be provided for these variables as well as for Creatine, INR, CK, Lipase, Amylase and Serum Creatinine.

A by-subject listing of abnormal laboratory values will also be presented, where the laboratory parameter name along with its out-of-range grade will be specified, for example, "Hemoglobin (low)," Hemoglobin (high)," etc. will be displayed in the listing.

9.6.4 Vital Signs

Vital signs (temperature, heart rate, respiratory rate, blood pressure recorded in the sitting position after at least 5 minutes rest) and weight will be assessed on all visits, i.e., Visit 1 (Week -2) through Visit 15 (Week 56), and ET Visit (if applicable). Height measurement will be performed only on Visit 1 (Week -2).

Observed and change from baseline in vital signs will be summarized by parameter and treatment group as defined in Section 5. Descriptive statistics will be shown for baseline, each post-baseline time point, and the change from baseline to each post-baseline time point.

9.6.5 Physical Examinations and Electrocardiogram (ECGs)

Complete physical examinations will be performed at screening (Visit 1 Week -2), Visit 7 (Week 8), Visit 14 (Week 52), Visit 15 (Week 56), and ET visit (if applicable). Symptom-directed (brief) examinations will be performed at Visit 2 (Day 1) through Visit 6 (Week 6), and at Visit 8 (Week 12) through Visit 13 (Week 39). Shift tables will summarize the worst post-baseline evaluation as compared to baseline visit by Treatment Group as defined in Section 5.

Clinically significant abnormal findings from brief physical examinations will be listed by subject.

A 12-lead ECG will be obtained in supine position after at least 5 minutes of rest at Visit 1 (Week - 2) through Visit 15 (Week 56), and ET Visit (if applicable). At Visit 2 through Visit 11 ECG will be performed

approximately 60-90 minutes after dosing. Shift tables will summarize the worst post-baseline evaluation as compared to baseline visit. A by-subject listing of abnormal ECG values will be provided.

9.7 Exploratory Analyses

The descriptive summaries of the exploratory variables listed in Section 4.1.3 Exploratory Variables will be provided for baseline, change and percent change from baseline at Week 12 and Week 52 by Initial dose dose groups and 5 mg cohort groups for the mITT population.

The correlation between clinical and biochemical/hematological biomarkers listed in Section 4.1.3 may be explored. For example, the correlation between total AP and bone-specific AP.

9.8 Subgroup Analyses

Subgroup analysis by cirrhotic status at baseline will be performed for the following endpoints:

- Efficacy endpoints: ALP, ALT, Total bilirubin, pruritus, VAS. The analyses will be performed by the initial treatment groups using the mITT population
- Safety: AE will be summarized by the initial treatment groups using the safety population.

10.0 Planned Differences from Protocol

10.1 Timing of Interim Analyses

The protocol states that an interim analysis for safety and efficacy will be performed after the 24th subject is enrolled. The first interim analysis was performed after the 24th subject performs their Week 8 visit. In addition, a second interim analysis was performed when the 24th subject completed their Week 12 visit. A third interim analysis was performed when the 45th subject completed their Week 26 visit. A fourth interim analysis was performed when the 40th subject completed their Week 52 visit.

10.2 Clinical Laboratory Analysis Baseline

The protocol states that two baselines will be used for the analysis of clinical laboratory assessments which are described as Baseline 1 and Baseline 2. This analysis plan uses the planned definition for Baseline 1 for all clinical laboratory assessments,

10.3 Study Completer

The protocol states that a subject must complete the Week 56 visit in order to be considered a study completer, however most subjects are expected to go directly from this study into the open label long term safety study at Week 52. Subjects who roll over into the long term safety study will be considered a completer in this study.

10.4 Analysis Treatment Group

The protocol states that analyses will be performed by the treatment group the subject was assigned at baseline. Additional treatment groups, 5 mg cohort groups, are included in the SAP.

10.5 Study Populations

The mITT population defined in the protocol is renamed as ITT population in the SAP. A revised mITT population is defined in the SAP as the primary population for efficacy analyses.

11.0 Validation

The programming (including quality control) of the analysis datasets and TFLs will be conducted by PRA Health Sciences (PRA). The programming will be conducted under PRA's standard operating procedure

PRS 050 and documented accordingly. The entire set of tables, figures and listings (TFLs) will be checked for completeness and consistency prior to its delivery to the Sponsor by the lead statistician and a senior level statistician, or above.

The PRA validation process will be repeated any time TFLs are redelivered using different data. Execution of this validation process will be documented through the study Table of Programs that will be provided to the client at study conclusion.

12.0 References

Hafner KB and Ruberg SJ (1996). Factorial dose response studies using frequency and magnitude of dose. Journal of Biopharmaceutical Statistics, 6(3), 253-262.

Elman, L.S. Hynan, V. Gabriel, and M.J. Mayo The 5-D itch scale: a new measure of pruritus. Br J Dermatol. 2010 162:587-93.

Appendix 1 Glossary of Abbreviations

7α-HC	7α-hydroxycholesterol
7-DHC	7-dehydrocholesterol
AE	Adverse Event
ALT	Alanine Aminotransferase
AMA	Anti-Mitochondrial Antibodies
ANCOVA	Analysis of Covariance
AP	Alkaline Phosphatase
Apo B-100	Apolipoprotein B-100
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Chemical
ATV	Atorvastatin
AUC	Area Under the Curve
β-SITO	β-sitosterol
BUN	Blood Urea Nitrogen
BQL	Below Quantifiable Limit
C4	7α-hydroxy-4-cholesten-3-one (bile acid precursor)
CA	Cholic Acid
CAMP	Campesterol
CD-1	Crl:CD1(ICR) Mice
CDCA	Chenodeoxycholic Acid
CK	Creatine Kinase
CK-MM	Creatine Kinase Muscle Type
Cmax	Maximum Plasma Concentration
COPR	Coprostanol
CRO	Contract Research Organization
CSTN	Cholestanol
CTX	C-terminal Telopeptide
CYP	Cytochrome P450
DCA	Deoxycholic Acid
DESM	Desmosterol
DSMB	Data Safety and Monitoring Board
EC	Ethics Committee
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
ELISA	Enzyme Linked Immunosorbent Assay

EOS	End of Study
EOT	End of Treatment
FDA	U.S. Food and Drug Administration
FGF19	Fibroblast Growth Factor 19
GCA	Glycocholic Acid
GCDCA	Glycochenodeoxycholic Acid
GCP	Good Clinical Practice
GDCA	Glycodeoxycholic Acid
GGT	Gamma-glutamyl Transferase
GLCA	Glycolithocholic Acid
GUDCA	Glycoursodeoxycholic Acid
HDL-C	High Density Lipoprotein Cholesterol
HoFH	Homozygous Familial Hypercholesterolemia
hs-CRP	High sensitivity C-reactive protein
IB	Investigator's Brochure
ICF	Informed Consent Form
ICH	International Conference on Harmonization
IgM	Immunoglobulin M
IXRS	Interactive Voice/Web Response System
LANO	Lanosterol
LATH	Lathosterol
LCA	Lithocholic Acid
LDH	Lactate Dehydrogenase
LDL-C	Low Density Lipoprotein Cholesterol
LOAEL	Lowest Observed Adverse Effect Level
LOCF	Last Observation Carried Forward
M	MBX-8025 Metabolite
MedRA	Medical Dictionary for Regulatory Activities
mITT	Modified Intent to Treat
NCI CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
OCA	Obeticholic Acid, Ocaliva
P1NP	Procollagen Type 1 N-terminal Propeptide
PBC	Primary Biliary Cholangitis/Primary Biliary Cirrhosis
PD	Pharmacodynamics
PK	Pharmacokinetics
PP	Per Protocol

PPARδ	Peroxisome Proliferator-activated Receptor Delta
PV	Pharmacovigilance
q.d.	Once Daily
QoL	Quality of Life
RBC	Red Blood Cells
SAE	Serious Adverse Event
SOP	Standard Operating Procedure
SQLN	Squalene
STIG	Stigmasterol
SUSAR	Suspected Unexpected Serious Adverse Reaction
t1/2	Half-life
TC	Total Cholesterol
TCA	Taurocholic Acid
TCDCA	Taurochenodeoxycholic Acid
TDCA	Taurodeoxycholic Acid
TEAE	Treatment Emergent Adverse Event
TG	Triglycerides
TLCA	Taurolithocholic Acid
Tmax	Time to Cmax
TUDCA	Tauroursodeoxycholic Acid
UDCA	Ursodeoxycholic acid
UK-PBC	United Kingdom Primary Biliary Cirrhosis Score
ULN	Upper Limit of Normal
UNS	Unscheduled
VAS	Visual Analog Score
VLDL-C	Very Low Density Lipoprotein Cholesterol
WBC	White Blood Cells
WHO	World Health Organization
WHOCC	World Health Organization Collaborating Centers

Appendix 2 List of Inferential Analysis Supportive SAS Output Appendices

TFL Appendix	Appendix Title

Appendix 3 Document History

Version Date	Modified/Reviewed By	Brief Summary of Changes (if created from a template, include template code)
03 Feb, 2017	[redacted]	Initial Creation
14 June, 2017	[redacted]	Updated for version 2 (dated 24 February, 2017) of the protocol including the following: • Treatment groups updated to 2, 5, 10, 20 mg • MicroRNA-122 endpoint added • Variables DK-MG and CK-BB added • Text around the interim analyses timing updated • PP population analyses updated to replicate only for the primary endpoint analyses
21 Dec, 2017	[redacted]	Updated for version 4.2 of the protocol (dated 20 July, 2017) including the following: Treatment groups updated to 2, 5, and 10 mg Study duration updated to 52 weeks with a 4 week FU period Primary analysis section updated to clearly denote Week 8 5 mg vs 10 mg % CFB on AP using the ANCOVA model as the primary analysis and other analyses being sensitivity/secondary analyses Linear trend test added Study design section updated with the figure from the new protocol and the schedule of events from the new protocol Baseline 2 removed. All analyses done on Baseline 1 Added definitions for Actual Treatment Group, Assigned Treatment Group, and Week 26 Treatment Group Added definition for Study Completers Added a third interim analysis
4 June 2019	[redacted]	Added the Rotterdam criteria and its definition, deleted first bullet of Toronto II Deleted change during extension period analysis and current max dose definitions Deleted on treatment definition Updated study period definitions for each period except follow-up to begin on Day 1 of the study in accordance with the analyses Updated the formula for the calculation of UK-PBC Risk Score to align with the Phase 3 study Added IA 4 (subjects at 52 Weeks with all other populations). Additional text added to the Sample Size Consideration section for the justification of sample size Removed text describing the assignment of subjects to 5mg and 10mg; and the registration of subjects to the 2mg treatment group based on computer generated randomization and registration list respectively Added "Model for End-Stage Liver Disease (MELD) scores" and "GLOBE score" to the Secondary Variables

		Added "for PK Analyses" to subsection title Deleted Current Max Dose definition Deleted Study Period Deleted Week 26 Endpoint UDCA compliance text was added under UCDA daily dose Added MELD and MELD-Na Score formula to the general deinitions section Added GLOBE Score formula to the general definitions Updated ITT population definition, and added mITT population Fourth Interim Analysis added Updated Total number of capsules planned formula Added "ITT and mITT populations" to PBC history summarization Added more text to specify populations used for efficacy analysis Estimate statements in PROC mixed syntax have been updated Deleted Secondary Analyses Baseline was added as a predictor in PROC mixed syntax "ANCOVA on change from Baseline" and Linear Trend Test for Dose Response" was removed "Subjects with Cirrhosis" was removed MELD score and GLOBE score were added under Secondary Analyses section Updates were made to Live Safety Monitoring Treatment-emergent fatal Aes was removed Texts about shift table summaries were added Texts about shift table summaries were added Subgroup analysis was added Additional reference added
26 November, 2019	[redacted]	Clarify per protocol definition Made correction in definition of PBC Response Criteria Clarify UDCA compliance definition