STATISTICAL ANALYSIS PLAN

Study Title:	Randomized, Double-Blind, Placebo-Controlled Phase 2 Study to Evaluate the Efficacy and Safety of Maralixibat in the Treatment of Subjects with Biliary Atresia after Hepatoportoenterostomy
Name of Test Drug:	Maralixibat
Study Number:	MRX-701
Protocol Version (Date):	Protocol Version 6 (25 April 2023)
Analysis Type:	Primary and Final Analysis
Analysis Plan Version:	1.0
Analysis Plan Date:	02 November 2023
Analysis Plan Author:	[redacted]

CONFIDENTIAL AND PROPRIETARY INFORMATION

Document History

Version	Date	Change(s)	Author
1.0	Click or tap to enter a date.	Initial version	[redacted]

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Definition
7αC4	7α-hydroxyl-4-cholesten-3-one
AE	adverse event
ALP	alkaline phosphatase
ALT	alanine aminotransferase
AST	aspartate transferase
APRI	AST-to-platelet ratio index
ATC	anatomical therapeutic chemical
BA	biliary atresia
BID	twice daily
BL	baseline
BMI	body mass index
CI	confidence interval
CMH	Cochran-Mantel-Haenszel
CRF	case report form
CSR	clinical study report
[redacted]	[redacted]
DMC	Data Monitoring Committee
ECG	electrocardiogram
EM	expectation-maximization
EMA	European Medicine Agency
EOT	end of treatment
ET	early termination
FDA	US Food and Drug Administration
[redacted]	[redacted]
GGT	gamma-glutamyl transpeptidase
HPE	hepatoportoenterostomy
ICH	International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
INR	international normalized ratio
IP	investigational product
IRT	interactive response technology
ITT	intent to treat
LLOQ	lower limit of quantitation
LS	least squares
LSV	lipid-soluble vitamin
MAR	missing at random
MCMC	Markov-Chain Monte-Carlo
MI	multiple imputation

Abbreviation	Definition
MMRM	mixed-effects model for repeated measurements
MNAR	missing not at random
NT	not treated
OLE	open-label extension
PDGP	protocol deviation guidance plan
[redacted]	[redacted]
PELD	Pediatric End-Stage Liver Disease
PG	propylene glycol
PK	pharmacokinetic
PP	Per protocol
PT	preferred term
Q1	first quartile
Q3	third quartile
QD	once daily
QTcB	QT interval corrected using Bazett's formula
QTcF	QT interval corrected using Fridericia's formula
REML	restricted maximum likelihood
RSS	Royal Statistical Society
SAE	serious adverse event
SAP	statistical analysis plan
sBA	serum bile acid
SD	standard deviation
SE	standard error
SI	International System of Units
SOC	system organ class
TEAE	treatment-emergent adverse event
TIP	target IBAT population
TLFs	Tables, Listings, and Figures
TSB	total serum bilirubin
UDCA	ursodeoxycholic acid
ULN	upper limit of normal
ULOQ	upper limit of quantitation
VBL	baseline visit
WHO	World Health Organization
WHO-DD	World Health Organization - Drug Dictionary

1. OVERVIEW

This statistical analysis plan (SAP) describes the planned analysis and reporting for Study MRX-701 (Randomized, Double-Blind, Placebo-Controlled Phase 2 Study to Evaluate the Efficacy and Safety of Maralixibat in the Treatment of Subjects with Biliary Atresia after Hepatoportoenterostomy), Protocol Version 6, dated 25-Apr-2023. Reference materials for this SAP include the protocol and the accompanying sample data collection documents. Operational aspects related to collection and timing of planned clinical assessments are not repeated in this SAP unless relevant to the planned analysis.

The structure and content of this SAP provide sufficient detail to meet the requirements identified by the Food and Drug Administration (FDA), European Medicines Agency (EMA), and International Conference on Harmonization (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use: Guidance on Statistical Principles in Clinical Trials (ICH 1998). All work planned and reported for this SAP will follow internationally accepted guidelines, published by the American Statistical Association (ASA 2018) and the Royal Statistical Society (RSS 2014), for statistical practice.

The planned analyses identified in this SAP may be included in clinical study reports (CSRs), regulatory submissions, or future manuscripts. Also, post hoc exploratory analyses not necessarily identified in this SAP may be performed to further examine study data. Any post hoc or unplanned exploratory analysis performed will be clearly identified as such in the final CSR.

The statistical plan described hereafter is an a priori plan. It will be submitted to file prior to any unblinded inferential or descriptive analysis of data pertaining to Study MRX-701.

2. STUDY OBJECTIVES AND ENDPOINTS

2.1. Double-Blind Period Study Objectives

2.1.1. Primary Objective

To evaluate the efficacy of maralixibat on biliary drainage after hepatoportoenterostomy (HPE) in participants with biliary atresia (BA).

2.1.2. Secondary Objectives

The secondary objectives are:

- To evaluate the rate of clinically relevant reductions in cholestatic biomarkers with maralixibat treatment
- To evaluate the rate of liver-related clinical events
- To evaluate the safety, tolerability, and pharmacokinetics of maralixibat

2.1.3. Exploratory Objectives

[redacted]

2.2. Open-Label Extension Period Study Objectives

To assess all of the above primary and secondary objectives analyzed over the full study duration, including the open-label extension (OLE) period.

2.3. Study Endpoints

2.3.1. Safety Endpoints

The safety, tolerability and pharmacokinetics endpoints of this study include the following:

- Incidence of treatment-emergent adverse events (TEAEs), including serious, related to study medication, leading to withdrawal, special interest TEAEs, along with TEAEs by severity and by relationship to study medication
- Change from baseline in safety laboratory, physical examination findings, vital signs, neurodevelopmental assessment, and maralixibat pharmacokinetic (PK) profile

Vital signs include heart rate, respiratory rate, body temperature, and blood pressure.

Safety laboratory tests and associated units of measure that will be used for reporting are listed in Appendix 1. Note that bilirubin (total and direct), alanine aminotransferase (ALT), and

gammaglutamyl-transpeptidase (GGT) are considered as both safety and efficacy laboratory tests.

2.3.2. Efficacy Endpoints

For the sequential (hierarchical) testing and the specific study populations used for the primary and each secondary endpoints in the double-blind phase, refer to Section 6.1.3 (Multiple Comparisons).

The primary and secondary endpoints are described below for the double-blind period of the study, but they will be assessed again for the OLE period of the study. The difference is that they will be assessed through Week 104 (for the OLE) instead of Week 26 (for the double-blind period).

2.3.2.1. Primary Efficacy Endpoint

The primary efficacy endpoint of this study is defined as the mean change in TSB levels from baseline through Week 26.

2.3.2.2. Secondary Efficacy Endpoints

The secondary efficacy endpoints of this study include the following:

- Mean change in total sBA levels from baseline through Week 26
- Proportion of participants with mean TSB levels <2 mg/dL through Week 26
- Proportion of participants observed to have a liver-related clinical event, including liver transplantation, liver decompensation (hepatic encephalopathy, variceal bleeding, new persistent ascites), discontinuations due to liver related events, or death through Week 26
- Proportion of participants undergoing liver transplantation or death through Week 26
- Proportion of participants observed to develop clinically evident portal hypertension defined as splenomegaly (spleen size >2 cm below the costal margin palpated on physical examination) and thrombocytopenia (platelet count <150 x 10^9/L) or clinically evident ascites or endoscopic evidence of esophageal or gastric varices through Week 26
- Proportion of participants with mean TSB levels ≤1.2 mg/dL through Week 26
- Proportion of participants with mean sBA levels ≤40 mmol/L through Week 26

2.3.2.3. Exploratory Efficacy Endpoints

[redacted]

[redacted]

3. OVERALL STUDY DESIGN AND PLAN

3.1. Overall Design

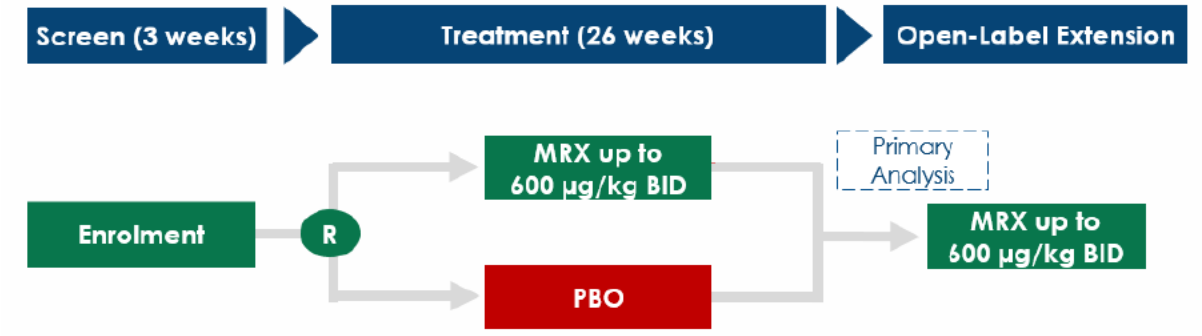
This is a 26-week, multicenter, double-blind, placebo-controlled, randomized parallel-group study, followed by an OLE that includes analyses up to Week 104, in participants with BA. In the double-blind period of the study, participants are randomized in a 1:1 fashion to receive either placebo or maralixibat—up to 600 µg/kg twice daily (BID; Figure 1). Treatment

assignment used a block randomization process scheme by study site (i.e., stratified by study site). The study periods are as follows:

1. Screening (1 week up to 3 weeks)
2. Double-blind period
    a. Dose escalation (4–8 weeks)
    b. Stable dosing (18–22 weeks)
3. OLE period (78 weeks)
    a. Dose escalation (4–8 weeks)
    b. Stable dosing (70–74 weeks)
4. Follow-up (2 weeks)

During the treatment period, participants receive standard-of-care treatment in line with investigator and caregiver preference and in addition to study medication.

Figure 1 Study Design



BID=twice daily; MRX=maralixibat; PBO=placebo; R=randomization.

All participants who complete the double-blind period enter the OLE, in which all participants receive open-label maralixibat at a maximum tolerated dose of 600 µg/kg BID. [redacted]

Participants remain in the OLE to at least study Week 104. [redacted]

For information about the schedule of assessments throughout the study, refer to Table 1 and Table 2 and to Section 7 in the protocol. For information on dosing, including dose-escalation requirements, refer to Section 6.2.3 in the protocol.

3.2. Sample Size and Power

Estimates for placebo response and variance are drawn from Bezerra et al. (2014). Under the assumption of no change over time in total bilirubin levels in the placebo group, a standard deviation of 3 mg/dL and a sample size of 33 participants per treatment arm provides 80% power to detect an expected reduction in total bilirubin compared with placebo of approximately 30% or 2.1 mg/dL by Week 26 of treatment, at the 2-sided alpha-level of 0.05, with use of a 1:1 randomization scheme. With allowance for a 10% attrition rate, 72 participants (36 per group) were planned for enrollment into the study.

3.3. Study Population

The study population comprises male and female participants with a body weight ≥2500 g (before or during the screening period) who are ≥21 days and <90 days of age at time of HPE or Kasai procedure. For a complete list of inclusion and exclusion criteria, refer to Sections 4.1 and 4.2 in the protocol, respectively.

3.4. Treatments Administered

Participants are randomized for the double-blind portion of the study to one of two treatment groups:

- Maralixibat
- Placebo

All participants who enter the OLE portion of the study receive open-label maralixibat.

3.5. Blinding and Unblinding

All participants, investigators, and study personnel involved in the conduct of the study, including data management, are blinded to treatment assignment.

The treatment assignment was not to be unblinded during the study except in emergency situations where the identification of the study drug is required for further treatment of the participant. The Investigator was to make an effort to contact the medical monitor before unblinding or as soon as possible after the Investigator has been unblinded without revealing the treatment assignment to the medical monitor. In any event, the medical monitor and Sponsor were to be informed about the code break as soon as possible.

If the treatment assignment was unblinded, the date and the signature of the person who was unblinded and the reason for unblinding was to be recorded in the source documents. Upon breaking the blind, the participant was to be withdrawn from the study but was to be followed for safety purposes. Unblinding was to be processed through the interactive response technology (IRT) system.

Data that may potentially unblind treatment assignment (i.e., maralixibat serum concentrations, treatment allocation, postbaseline sBA levels) are to be handled with special care during the data

cleaning and review process. Prior to unblinding, any data that may potentially unblind study site personnel or study team personnel are to be presented as blinded information or otherwise will not be made available. If applicable, unblinded data were to be made available to quality assurance representatives for the purposes of conducting independent audits.

3.6. Schedule of Assessments

A detailed schedule of assessment for the study, as presented in the protocol, is provided in Table 1 for the double-blind period and in Table 2 for the OLE period.

Table 1 Schedule of Assessments: Screening and Double-Blind Period

Procedure	Screening[a]	BL	Dose Escalation (Duration: 4–8 Weeks)[b,c,d]					Stable Dosing (Duration: 18–22 Weeks)						
Visit/Participant Contact	VS	VBL	[redacted]											
Study Week	-3 to -1	0												
Study Day	-21 to -1	0												
Window (days)														
Informed consent	X													
Eligibility assessment	X	X												
Demographics	X													
Medical history	X													
Breastfeeding status	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Physical examination, vital signs[g]	X	X	X	X		X		X		X		X		X
[redacted]		X						X						X
[redacted]		X						X						X
[redacted]		X[j]												X
[redacted]		X												X
Electrocardiogram		X						X						X
[redacted]		X						X						X
[redacted]		X	X	X		X		X		X		X		X
CBC with differential[m]	X			X				X				X		X
Coagulation[m]	X					X		X		X		X		X
Chemistry panel[n]	X	X		X		X		X		X		X		X
sBA collection[o, p]		X		X		X				X		X		X
Retinol/Vitamin A, α-tocopherol/Vitamin E [p, q]	X					X				X				
25-hydroxy vitamin D[p, q]	X							X				X		X
[redacted][o]		X		X		X				X		X		X
PK sample[r, o]		X		X				X						X
[redacted]		X								X				X
[redacted]	X													
Study medication supplied[s]		X	X	X		X		X		X		X		
Study drug administration		X	X	X	X	X	X	X	X	X	X	X	X	X
Assess AEs	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Prior/concomitant treatments	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Assess participant dosing compliance[t]			X	X	X	X	X	X	X	X	X	X	X	X

[redacted]; AE=adverse event; BL=baseline; CBC=complete blood count; DMC=Data Monitoring Committee; [redacted] PK=pharmacokinetic; sBA=serum bile acid; V=visit.

a Participants who initially do not meet eligibility criteria may be reassessed during the 3-week screening period prior to being considered as a screen failure. [redacted].

b During the dose-escalation period, dose escalation should occur in the absence of major safety (e.g., liver parameters) or tolerability concerns (e.g., gastrointestinal-related treatment-emergent adverse events) related or possibly related to the study medication. [redacted]

c Additional study visits at the investigator's discretion.

d Dose escalation visit schedule may be updated per DMC recommendation [redacted] participants have completed the dose-escalation period of the double-blind period, per Section 6.2.4.1 in the Protocol.

e [redacted]

f Assessments at Week 26 of the double-blind period are combined with the assessments of the Week-26 OLE period at the same study visit. All assessments/measurements from this double-blind dosing period visit to also apply for the first visit of the open-label extension period.

g Length, weight, blood pressure, heart rate, temperature, and respiration rate. Spleen size below the lower costal margin (measured in centimeters) will be assessed and recorded by the investigator.

h [redacted]

i [redacted]

j Should be performed within a [redacted] of the BL visit.

k [redacted]

l [redacted]

m CBC and coagulation results performed as part of standard of care by the local laboratory may be used for the study if sample was collected within 3 days before or after the study visit.

n Clinical chemistry (including total and conjugated bilirubin, ALP, ALT, AST, GGT, and albumin) and sBA samples must always be sent to the central laboratory.

o The sBA, [redacted], and PK results will remain blinded to sites and to the blinded study team (see Section 6.1.2 in the Protocol for exceptional circumstances).

p Samples should be taken before feeding, with a suggested fasting of approximately 2 hours prior to collection and before administration of vitamin supplementation, when possible. Water intake, excluding milk, is permitted if necessary.

q Lipid-soluble vitamin values performed as part of standard of care by the local laboratory may be used for the study if samples were collected within ±14 days from the study visit. If blood volumes do not allow all assessments within a single visit, a separate clinical visit in which samples for retinol/Vitamin A, α-tocopherol/Vitamin E, and 25-hydroxy vitamin D are taken can be made.

r Blood samples should be collected before feedings and administration of vitamin supplementation, when possible. Systemic concentrations of maralixibat in plasma will be determined at predose and at approximately [redacted]. Predose samples are not required for measures made at BL.

s If needed, and where permitted, study medication may be supplied via direct shipment to participants between site visits.

t [redacted] compliance is assessed at each visit.

Table 2 Schedule of Assessments: Open-Label Extension and Follow-Up Periods

Procedure	Dose Escalation (Duration: 4–8 Weeks)[a,b,c]					OLE Stable Dosing (To Week 104)										OLE Stable Dosing (Beyond Week 104)[d]		ET[e] /EOT	Follow-Up after ET/EOT
Visit/Participant Contact	[redacted]																		
Study Week																			
Study Day																			
Window (days)																			
Breastfeeding status	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Physical examination, vital signs [h]		X	X		X		X		X		X		X		X	X		X	X
[redacted]							X		X		X		X		X	X		X	X
							X		X		X		X		X	X		X	X
Electrocardiogram																X			X[j]
[redacted][k]			X						X		X		X		X	X		X	X
		X	X		X		X		X		X		X		X	X		X	X
										X						X			
																			X
CBC with differential [n]					X		X		X		X		X		X	X		X	X
Coagulation [n]					X		X		X		X		X		X	X		X	X
Chemistry panel [o]			X		X		X		X		X		X		X	X		X	X
sBA collection [p,q]			X		X		X		X		X		X		X	X		X	X
Retinol/Vitamin A, α-tocopherol/Vitamin E [q, r]			X				X		X		X		X		X	X		X	X
25-hydroxy vitamin D [q, r]					X		X		X		X		X		X	X		X	X
[redacted]			X		X		X		X		X		X		X	X		X	X
										X						X			X
Study medication supplied [s]	X		X		X		X		X		X		X		X	X		X	
Study medication administration		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Assess AEs	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Prior and concomitant treatments		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Assess participant dosing compliance[t]		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	

[redacted] AE=adverse event; CBC=complete blood count; circ=circumference; DMC=Data Monitoring Committee; EOT=end of treatment; ET=early termination; OLE=open-label extension; [redacted] sBA=serum bile acid; V=visit.

a During the dose-escalation period, dose escalation should occur in the absence of major safety (e.g., liver parameters) or tolerability concerns (e.g., gastrointestinal-related treatment-emergent adverse events) related or possibly related to the study medication. [redacted]

b Additional study visits at the investigator's discretion.

c Dose escalation visit schedule may be updated per DMC recommendation [redacted] participants have completed the dose-escalation period of the double-blind period, per Section 6.2.4.1. in the protocol.

d Study medication to continue twice daily in the setting of treatment continuation. [redacted]

e If the ET/EOT visit is within 7 days of the last repeating visit, blood and other clinical assessments do not need to be repeated if previously performed unless there is clinical suspicion of abnormality. Breastfeeding status, AEs, participant dosing compliance, and prior and concomitant treatments will be assessed regardless of timing of the last repeating visit.

f Assessments at Week 26 of the OLE period are combined with the assessments of the Week-26 double-blind period at the same study visit. All assessments/measurements from this double-blind visit to also apply for the first visit of the OLE period.

g [redacted]

h Length, weight, blood pressure, heart rate, temperature, and respiration rate. Spleen size below the lower costal margin (measured in centimeters) will be assessed and recorded by the investigator.

i [redacted]

j ECG is not required at ET/EOT if the ET/EOT visit occurs after Week 104.

k [redacted]

l [redacted]

m [redacted]

n CBC and coagulation results performed as part of standard of care by the local laboratory may be used for the study if sample was collected within 3 days before or after the study visit.

o Clinical chemistry (including total and conjugated bilirubin, ALP, ALT, AST, GGT, and albumin) and sBA samples must be sent to the central laboratory.

p The sBA [redacted] results will remain blinded to sites and to the blinded study team (see Section 6.1.2 in the Protocol for exceptional circumstances).

q Samples should be taken before feeding, with a suggested fasting of approximately 2 hours prior to collection and before administration of vitamin supplementation, when possible. Water intake, excluding milk, is permitted if necessary.

r Lipid-soluble vitamin values performed as part of standard of care by the local laboratory may be used for the study if samples were collected within ±14 days from the study visit. If blood volumes do not allow all assessments to be performed within a single visit, a separate clinical visit in which samples for retinol/Vitamin A, α-tocopherol/Vitamin E, and 25-hydroxy vitamin D are taken can be made.

s If needed, study medication may be supplied via direct shipment to participants between site visits.

t [redacted]

4. STATISTICAL ANALYSES AND REPORTING

4.1. Introduction

Data processing, tabulation of descriptive statistics, calculation of inferential statistics, and graphical representations will be performed p rimarily using SAS (release 9.4 or higher).

Continuous (quantitative) variable summaries will include the number of participants with nonmissing values (n), mean, standard deviation (SD) and/or standard error (SE) if appropriate, median, Q1, Q3, minimum, and maximum.

Categorical (qualitative) variable summaries will include the frequency and percentage of participants who are in the particular category for each possible value. In general, the denominator for the percentage calculation will be based upon the total number of participants in each treatment group with available data in the analysis population, unless otherwise specified.

The minimum and maximum will be reported with the same degree of precision (i.e., the same number of decimal places) as the observed data. Measures of location (mean and median) will be reported to 1 degree of precision more than the observed data, and measures of spread (i.e., SD or SE) will be reported to 2 degrees of precision more than the observed data. Confidence intervals (CIs) are reported to the same degree of precision as parameter they are calculated for.

The minimum and maximum values for derived and select observed values will be reported as follows, with measures of location and spread following the above rules. Derived values for corrected sodium (mEq/L) and PELD scores will be presented as integers. Derived values for [redacted] summary and total scale scores will be presented to 1 decimal place. sBA levels (total and subspecies) and derived values of height, weight, and BMI z-scores, APRI, retinol:RBP molar ratio (mol/mol), and ratio of αtocopherol to the sum of cholesterol and triglycerides (mg/g) will be reported to 2 decimal places.

Percentages will be presented to 1 decimal place, unless otherwise specified. Where the number of participants in a particular category is zero, a percentage (i.e., 0.0%) will not be displayed.

Unless otherwise indicated, all testing of statistical significance will be 2-sided, and a difference resulting in a p-value of <0.05 will be considered statistically significant. Corresponding 95% CIs will be presented for statistical tests.

A p-value of ≤ 0.10 but ≥ 0.05 will be considered evidence of a trend.

To control for the overall type I error rate, hierarchical testing based on a fixed sequence procedure will be used (see Section 6.1.3). If statistical significance is declared for the primary efficacy analysis, formal hypothesis testing will be done for the secondary efficacy endpoints in the prespecified sequence until a nonsignificant result is reached. All other p-values from secondary endpoints, after a nonsignificant p-value is reached, will be considered nominal and will be interpreted with appropriate caution when considering the extent to which results from secondary endpoints provide support for evidence of efficacy.

By-participant listings will be presented for all participants in the Safety Analysis Set and sorted by participants identification (ID) number, visit date, and time (if applicable). Data collected on log forms, such as adverse events (AEs), will be presented in chronological order for each participant. Age, sex, race, and ethnicity will be included in the listings, as space permits.

4.2. Interim Analysis and Data Monitoring

No interim analyses are planned.

A DMC will be involved in the management of this study. DMC meetings will be held periodically for the duration of the study. The purpose of the DMC is to review the progress of the study, with regard to safety, and to make recommendations to stop or modify the study if safety concerns are identified. In addition to scheduled meetings, the DMC will convene for ad hoc meetings in case of any safety concerns arising during the conduct of the study.

Further details regarding the DMC can be found in the DMC Charter, which was available prior to the enrollment of the first participant. There is no study stop planned based on the efficacy results during the study, and no alpha spending will occur prior to the Primary Analysis.

4.3. Primary Analysis

A single and formal Primary Analysis will be performed after the last participant has completed Week 26 or prematurely discontinued the study.

4.4. Final Analysis

The final analysis will be performed after all participants have completed or prematurely discontinued the study and will focus on the Week 104 endpoints from the OLE.

5. ANALYSIS POPULATIONS

The following analysis populations are planned for this study:

- **Safety Population:** The Safety Population will consist of all participants who receive at least 1 dose of study drug.
- **Intent-To-Treat (ITT) Population:** The ITT Population will consist of all randomized participants.
- **Per-Protocol (PP) Population**: The PP Population will consist of all participants in the ITT Population who receive at least 1 dose of study drug and do not have any important protocol violations or deviations that have a potential impact on the efficacy analysis. Important protocol violations/deviations will be identified prior to database lock.
- **Target IBAT Population:** The target IBAT population (TIP) will consist of participants who do not normalize total bilirubin in the early stages of the study. The criteria for selection into this population will be determined based on the

placebo-treated participants and natural history data and then applied to the maralixibat-treated participants.

Membership in the analysis populations will be determined prior to database lock except for the TIP, which requires unblinded data by its definition.

6. GENERAL CONSIDERATIONS FOR STATISTICAL ANALYSES

6.1. Statistical Definitions and Algorithms

6.1.1. Baseline

In general, baseline values are defined as data measured or collected at the baseline visit, VBL (Day 0), before the first dose of study drug is administered. If data are not measured or collected at the baseline visit, then the last nonmissing measurement before receiving the first dose of study drug is used as the baseline value.

Baseline and all postbaseline sBA samples are assayed at Frontage Laboratories.

For liver-biochemistry laboratory parameters (i.e., ALT, AST, ALP, total bilirubin, direct bilirubin, GGT, and albumin) baseline is defined as the average of data collected post-Kasai between and including Screening and VBL (Day 0), or the last nonmissing measurement before receiving the first dose of study drug if only one value is available. Predose measurements obtained from a local laboratory will be used only if no other data are available. For the derived noninvasive fibrosis markers of APRI and PELD, the same (derived/average) liver-biochemistry baseline value will be used.

6.1.2. Adjustments for Covariates

Efficacy variables that are continuous measures assessed over time will be analyzed using a mixed-effects model for repeated measures (MMRM) with change from baseline as the dependent variable and fixed, categorical effects of site, treatment group, analysis visit, and treatment group-by-visit interaction as well as the continuous, fixed covariates of baseline and baseline-by-visit interaction.

No other covariates will be included in the analyses of the efficacy endpoints.

6.1.3. Multiple Comparisons

The type I error rate of $\alpha=0.05$ will be maintained for the study by using sequential (hierarchical) testing for primary and secondary efficacy endpoints.

The testing will be done in the following order:

1. Primary: mean change in TSB from baseline through Week 26
2. Secondary: proportion of participants with TSB levels <2 mg/dL through Week 26
3. Secondary: mean change in total sBA levels from baseline through Week 26

4. Secondary: proportion of participants observed to have a liver-related clinical event, including liver transplantation, liver decompensation (hepatic encephalopathy, variceal bleeding, new persistent ascites), or death through Week 26
5. Secondary: proportion of participants undergoing liver transplantation or death through Week 26
6. Secondary: proportion of participants observed to develop clinically evident portal hypertension defined as splenomegaly (spleen size >2 cm below the costal margin palpated on physical examination) and thrombocytopenia (platelet count <150 x 10^9/L) or clinically evident ascites or endoscopic evidence of esophageal or gastric varices through Week 26
7. Secondary: proportion of participants with TSB levels ≤1.2 mg/dL at Week 26
8. Secondary: proportion of participants with sBA levels ≤40 mmol/L at Week 26

Because the testing is sequential, the type I error rate of α=0.05 is maintained. Failure at any stage in the sequence implies no type I error control for the additional subsequent tests. All p-values for each comparison without adjustments will be provided in the summary tables for informational purposes.

6.1.4. Handling of Dropouts or Missing Data

Although all possible efforts will be made to ensure that participants stay in the study and that all data are collected as scheduled, the occurrence of missing data cannot be completely eliminated.

The primary analysis population for all efficacy analyses will be the ITT Population. For the analysis of primary efficacy endpoint and the secondary efficacy endpoint of mean change in total sBA, several methods will be used to handle missing data for TSB, including:

- MMRM
- multiple imputation (MI) using standard missing at random (MAR) approach
- MI using tipping-point approach

The MMRM method will be used as the primary analysis method on all continuous, repeated efficacy measures. Sensitivity analyses using the 2 MI methods will be performed on the primary efficacy endpoint to assess the robustness of alternate imputation assumptions.

For the secondary endpoint proportion of TSB level < 2 mg/dL, the MI generated for the primary endpoint will be used to impute the through Week 26.

The above procedures are described in the below subsections, along with the handling of missing individual laboratory data and AE severity and relationship to study drug.

6.1.4.1. MMRM Analysis Method

Efficacy variables that are continuous measures assessed over time will be analyzed using a MMRM model as the primary analysis method, with change from baseline as the dependent variable and fixed, categorical effects of site, treatment group, visit, and treatment group-by-visit interaction as well as the continuous, fixed covariates of baseline and baseline-by-visit interaction. Site is used as a surrogate for HPE surgeon.

The MMRM method has been demonstrated extensively as an appropriate choice for the primary analysis in longitudinal confirmatory clinical trials with continuous endpoints (Mallinckrodt et al. 2008). This analysis method, which is from a broader class of direct-likelihood analyses methods, makes use of fully and partially observed data sequences from individual participants by estimating the covariance between data from different time points (Molenberghs and Kenward 2007). Further, it is often useful to implement MMRM using an unstructured approach to modeling both the treatment-by-time means and the variances and covariances, leading to what is essentially a multivariate normal model wherein treatment group means at the primary time point are adjusted to reflect both the actual observed data and the projected outcomes from the participants with missing data (Cnaan et al. 1997; Molenberghs et al. 2004; Molenberghs and Kenward 2007).

As a direct likelihood method, the MMRM method is a preferred approach for handling missing data in such designs and will be used as the primary analysis method for all efficacy endpoints where assessments are made over time. MMRM is a full multivariate model in nature, which avoids potential bias as a predetermined model and operates in a more general MAR framework (Mallinckrodt et al. 2001). Data are considered MAR if, conditional upon the independent variables in the analytic model, the missingness depends on the observed outcomes of the variable being analyzed but does not depend on the unobserved outcomes of the variable being analyzed. This assumption implies that the behavior of the post-dropout observations can be predicted from the observed variables and therefore that treatment effect can be estimated without bias using the observed data (EMA 2010). For studies of missing data in a controlled clinical trial setting, MAR is usually considered as a plausible underlying missing mechanism (Molenberghs and Kenward 2007; Mallinckrodt et al. 2008, 2013; Siddiqui et al, 2009). The assumption of MAR is often reasonable because, particularly in longitudinal studies wherein the evolution of treatment effects is assessed by design over time, the observed data and the models used to analyze them can explain much of the missingness (Little et al. 1987; Verbeke and Molenberghs 2000). This point may be especially relevant in well-controlled studies, in which extensive efforts are made to observe all outcomes and factors that influence them while participants are following protocol-defined procedures. Thus, longitudinal clinical trials by their very design aim to reduce the amount of missing not at random (MNAR) data (missingness explained by unobserved responses), thereby increasing the plausibility of MAR (Mallinckrodt et al. 2008).

Code similar to the following SAS pseudocode will be used for the MMRM models:

6.1.4.2. Multiple Imputation Methods

Although the assumption of MAR, as used for the primary analysis method, is often reasonable in clinical trials, the possibility of MNAR data cannot be ruled out. Therefore, an analysis valid under MNAR will also be performed. Both MNAR- and MAR-based analyses using MI methods will be the basis upon which sensitivity of the analysis to missing data is assessed.

Any participant who withdraws or is discontinued from the study or who misses a scheduled visit or assessment up through Week 26 will have his or her primary and secondary efficacy of mean change in total sBA missing data analyzed as imputed using MI techniques. This analysis will be presented as a sensitivity analysis.

Multiple imputation is a simulation-based approach where missing values are replaced using an appropriate stochastic model given the observed data and covariates, creating multiple completed data sets. These completed datasets are then analyzed using standard analysis methods (MMRM for this study), and the different parameter estimates across the datasets are then combined to produce unique point estimates, SEs, and CIs with the uncertainty of the imputation process taken into account.

In most randomized clinical trials that collect data over time, the great majority of missing data follow a monotone pattern. That is, once a participant has missing data for some visit, data will be missing for all subsequent visits. Typically, there is also a small amount of nonmonotone missing data (i.e., some participants have missing values for intermediate visits, but have nonmissing data at subsequent visits).

The following 2 MI analysis models, one based on the standard MAR approach and the other based on the MNAR approach, will be used to examine robustness of the primary analysis results.

Standard MAR Imputation

The MAR imputation model will impute missing values using a regression-based multiple imputation model (Little and Yau 1996). For participants with complete data up to a particular timepoint (e.g., laboratory results for the primary endpoint and the secondary endpoint of mean change in total sBA), a multiple regression model will be fit that includes the outcome at that time point as the dependent variable and observed data (e.g., outcomes at previous time points, treatment, and baseline) as independent variables. Separate models will be similarly constructed for each time point. With use of these regression models, a missing value for a participant at a particular time point will be imputed as a draw from the predictive distribution given the outcomes at previous time points (some possibly imputed), treatment group, etc. This process will be repeated a given number of times (as specified below), resulting in the same number of complete analysis data sets. The MMRM analyses will be performed separately for each of the completed analysis data sets, and the results will be combined into one multiple imputation inference (Little and Yau 1996; Schafer 1997). This strategy is appropriate for data sets that have a monotone missing data pattern. If the data set does not precisely have this pattern, a monotone data augmentation method, as described below in Step 1, will be used to impute the small amount of missing data that is required to make the missing data pattern monotone before applying the MI algorithm described above.

Multiple imputation based on a standard MAR imputation approach will be performed in SAS with use of a general three-step approach:

Step 1: If the data have a non-monotone pattern of missingness, then a monotone data augmentation method using Markov-Chain Monte-Carlo (MCMC) will be used to impute data that are missing and required to make the missing data pattern monotone. Twenty datasets with a monotonic missing pattern will be generated. This method will use a noninformative Jeffreys prior to derive the posterior mode from the expectation-maximization (EM) algorithm as the starting values for the MCMC method. Intermittent missing values will be imputed using the MCMC method assuming a multivariate normal distribution over all variables included in the imputation model (i.e., treatment group, baseline, and each postbaseline visit). The MCMC statement of the MI procedure in SAS (PROC MI) will specify the CHAIN=MULTIPLE option, so that the procedure uses multiple chains and completes the default 200 burn-in iterations before each imputation, and the IMPUTE=MONOTONE option to create the 20 partially imputed datasets with a monotone missing pattern. The seed of the pseudorandom number generator used to randomly generate imputations for the missing values in Step 1 is [redacted]

Assumptions underlying the partial imputation step are such that participants with missing data follow the same model as other participants in their respective treatment arm that have complete data.

If the raw data have a monotone pattern of missingness, then the same procedures described above can be followed to create 20 identical datasets that will be used as an input dataset for the next step.

Pseudocode for imputation of primary efficacy missing data is provided below:

[redacted]

Step 2: Using a standard MAR-based MI approach, the remaining missing data will be imputed using a method for monotone missingness. The dataset that contains the multiple (20) partially imputed datasets is first sorted by imputation number and treatment group. The MI procedure used to complete the imputation will use a BY imputation number statement and request one imputed dataset within each BY group. The variables in the imputation model include treatment group, baseline, and each postbaseline visit. The final 20 fully imputed datasets will be generated using a regression-based multiple imputation model (PROC MI statement MONOTONE REGRESSION). For participants with complete data up to a particular visit, a multiple regression model will be fit that includes the outcome at that visit as the dependent variable and outcomes at previous visits and treatment group as independent variables. With use of these regression models, a missing value for a participant at a particular time point will be imputed as a draw from the predictive distribution given the outcomes at previous time points (some possibly imputed) and treatment group. The seed number of [redacted] will be used for the imputation procedure described in Step 2.

For both Steps 1 and 2, minimum and maximum values for TSB (i.e., 0 and 39xULN) will be specified in the MI procedure to avoid imputed values outside the possible range of values. When an intended imputed value is less than the minimum or greater than the maximum value specified, the MI procedure in SAS will redraw another value for imputation.

Pseudocode for this step is below:

[redacted]

Step 3: MMRM analyses will be performed separately for each of the 20 complete analysis datasets, and the results will be combined into one multiple imputation inference (estimated treatment effect, SE, p-value and associated 95% CI) using the SAS MIANALYZE procedure. The treatment difference will be tested at the 2-sided 0.05 level, and corresponding 95% CIs will be calculated. In the case that there are no missing data for a particular visit, p-values and 95% CIs will come from the MMRM analysis on the observed data.

Sample SAS pseudocode for the MMRM analysis is as follows:

```
ameterestimates=lsmn_2;
```

Tipping-Point Analysis

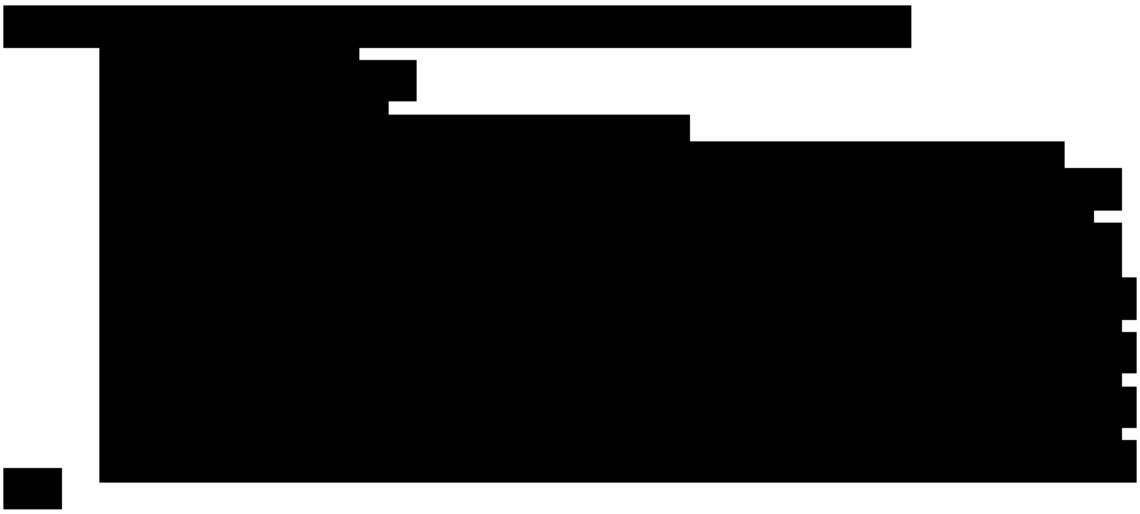
As a sensitivity analysis to the standard MAR imputation approach, a tipping-point analysis will be performed in order to determine the inflection point at which the inference under the MNAR assumption changes substantially. This will be used to check the robustness of the imputation.

The sensitivity analysis will be performed by using a specified sequence of shift parameters, which will adjust the imputed values for observations in each treatment group. The range of shift parameters to be included in this analysis are -4 to 4 by 0.5. Thus, the value at which the results of the analysis are shifted from significant (i.e., $\alpha<0.05$) to nonsignificant (i.e., $\alpha\geq0.05$) will be determined.

A heatmap figure will be generated to display p-values corresponding to all combinations of shifts from the imputed data.

The seed number of 3215487 will be used for the imputation procedure described in Step 2.

Steps 1 and 3 of the analysis will be the same as for the multiple imputation analysis as described in the Standard MAR Imputation above. However, Step 2 of the analysis is the step that the shift parameters will be applied. Pseudocode for Step 2 is as follows:



YY will encompass the range of shift parameters as prespecified above.

6.1.4.3. Missing [redacted]

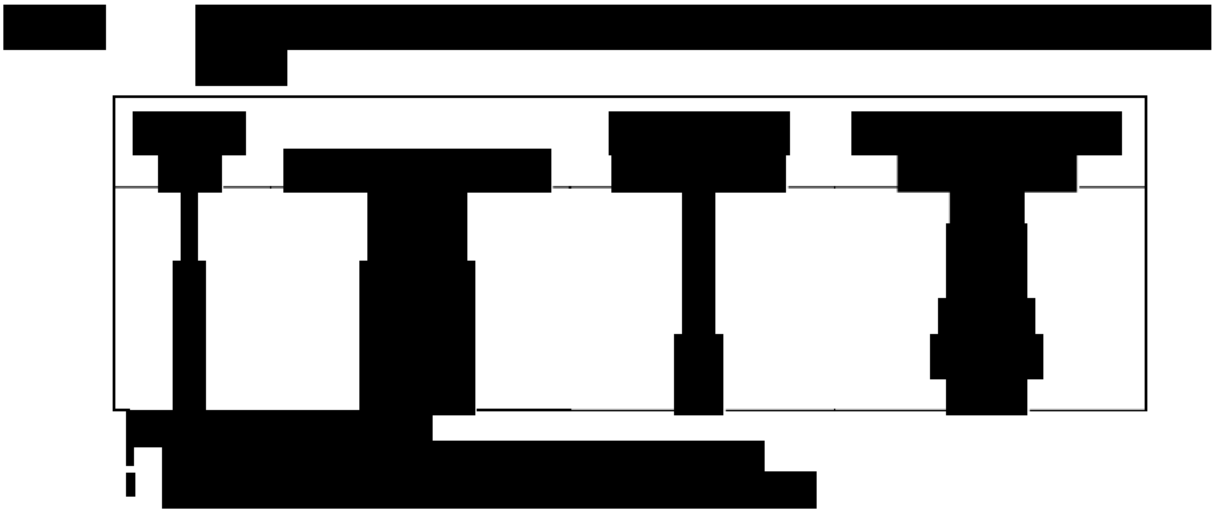
For [redacted], if >50% of the items in the scale are missing, the scale score is not computed.

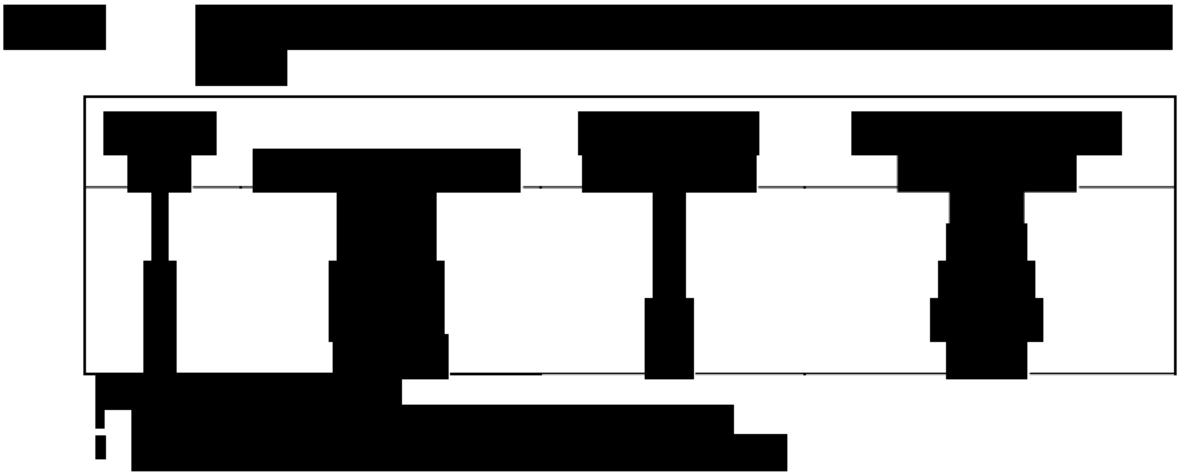
6.1.4.4. Missing Adverse Event Severity/Relationship

For analysis purposes, the following rules will be applied for missing AE severity or relationship to study drug. An AE that does not have a recorded relationship to study drug value will be conservatively considered as "Related" to study drug. If the severity of an AE is missing, the severity will be reported as "Missing."

6.1.5. Analysis Visit Windows

Analyses of all visit-based efficacy and safety variables will be performed using the analysis visit windows defined by study day relative to the first dose of study drug as outlined in Table 3 and Table 4. Scheduled visits will be selected over unscheduled visits. The below table addresses scheduled postbaseline in-clinic visits; baseline assessments are described in Section 6.1.1.





For those participants who discontinue early from the study, Table 3 and Table 4 will also be used to assign the appropriate analysis visit.

The study day will be calculated for each scheduled or ET postbaseline visit and compared with the assessment window presented in Table 3 to define the visit window used for analyses. The analysis visit windows apply only to those visits that are applicable to the specific assessment. For example, if the scheduled or ET visit falls at Week 10 but a specific assessment (e.g., sBA sample) was not scheduled at that visit (see Table 1, Schedule of Assessments), then that assessment will not be used for analyses at that particular visit.

If >1 assessment occurs within a single visit window, then the analysis will use the assessment closest to the target day. If 2 assessments within the same visit window are equidistant from the target day, then the analysis will use the later assessment.

Efficacy assessments performed >7 days after study drug was stopped will be excluded from efficacy analyses.

6.1.6. Investigative Sites

An investigative site is defined as a single principal investigator (including subinvestigators) who enrolls participants for the study. If an investigator has multiple practice locations, these locations are considered a single investigative site.

There is the potential that a participant could be transferred to a principal investigator who did not enroll the participant. Unless otherwise specified, the investigative site of the enrolling investigator will be used for the unique participant ID.

The primary presentations and analyses will be based on data pooled across investigative sites.

6.1.7. Treatment Period Definitions

- "Dose Escalation Period" is defined for analysis purposes as the period of time between the first dose of study drug and the end of study Week 8 (i.e., through the Week 8 visit date).

- "Stable Dosing Period" is defined for analysis purposes as the period of time between the beginning of Week 9 (i.e., Week 8 participant contact date plus 1 day) and the end-of-treatment (EOT) visit (inclusive).

6.1.8. Derived Variables

Select derived variables will be rounded for presentation purposes (see Section 4.1).

- **Age (days) at baseline =** date of first dose – date of birth + 1
- **Date of Birth:** The reported date of birth will be used.
- **UDCA usage at baseline**: A participant with reported use of ursodeoxycholic acid (UDCA), based on ATC Class Level 5 (chemical substance), with a start date on or prior to the first dose of study drug and either ongoing or with a stop date on or after the first dose of study drug would be considered as using UDCA at baseline.

- **Time since original diagnosis of BA (days)** = (date of first dose – date of original diagnosis of BA)

- **Time since HPE (days)** = (date of first dose – date of HPE)

- **Change from baseline** = postbaseline value at time point – value at baseline
- **% Change from baseline** = 100 x change from baseline / value at baseline

- **APRI** = 100 x $\frac{AST\left(\frac{U}{L}\right) / AST\ ULN\left(\frac{U}{L}\right)}{platelet\ count\left(\frac{10^3}{\mu L}\right)}$, *where ULN = upper limit of normal*

- **Ratio of Alpha Tocopherol to the sum of Cholesterol and Triglycerides (mg/g)** = 1000 x alpha tocopherol (mg/dL) / [cholesterol (mg/dL) + triglycerides (mg/dL)]
  For alpha tocopherol concentrations reported as below the minimum quantitation limit, half of the minimum quantitation limit is used in the calculation.

- **Corrected Sodium (mEq/L)** = sodium (mEq/L) + (0.002 x triglycerides [mg/dL])

- **PELD Scores**

  PELD score will be calculated for children under 12 years of age at the baseline visit.

  PELD Score = 4.80 × ln(total bilirubin [mg/dL]) + 18.57 × ln(INR) – 6.87 × ln(albumin [g/dL]) + 4.36 (if patient <1 year: scores for patients listed for liver transplantation before the patient's first birthday continue to include the value assigned for age (<1 year) until the patient reaches the age of 24 months) + 6.67 (if the patient has growth failure, where growth failure is defined as a weight z-score < -2.00)

  Laboratory values in the PELD equation that are <1.0 will be set to 1.0 for the calculation of the PELD score.

  Scores from PELD will be summarized.

- **Treatment Duration (days)** =

  Date of last dose of study drug – Date of first dose of study drug + 1 day – Prescribed dose interruptions

  For participants who are missing the date of last dose of study drug, the last known contact date will be used in the calculation of treatment duration. Prescribed dose interruptions are captured as gaps in dates between log lines in the study drug administration page.

- **Study Drug Exposure (days)** =

  Treatment duration (days) – Number of days that a participant reported missing both morning and evening doses (between the date of first and last dose)

  Study drug dosing errors are captured in the Study Drug Compliance page.

- **Compliance (%)** = 100 x Study drug exposure (days) / Treatment duration (days)

  Study drug compliance will not be calculated for participants whose date of last study drug is unknown.

- **Total Dose (µg/kg)** = $\sum$ [Number of doses taken$_i$ x Dose received (µg/kg)$_i$]

  where

  i = 1 to k, (k = number of time periods participant is receiving a constant dose)

- **Average Daily Dose (µg/kg/day)** = Total Dose (µg/kg) / Treatment Duration (days)

  Dose variables (total dose and average daily dose)will not be calculated for placebo participants.

- [redacted]

- **Treatment-Emergent Adverse Event** = In general, TEAEs are defined as AEs that start or deteriorate on or after the first dose of study drug and no later than 7 days following the last dose of study drug (for those not participating in the extension study). For participants with >7 days of study drug interruption/withdrawal, the definition of a TEAE considers both the date of the last dose prior to drug interruption and the actual last dose (see Section 6.1.9).

  Any event that started before the first dose and worsens in either intensity or frequency or changes from nonserious to serious on or after the first dose date will also be designated as a TEAE.

  For any participants who die during the study and the date of death is between the date of first dose of study drug and the date of study discontinuation (as entered by the site), inclusive, all AEs (including those resulting in death) that occur during the study will be considered as TEAEs irrespective of the last dose and will be included in the TEAE summaries.

- **Weight, Height and BMI Z-Scores**
  Height, weight, and body mass index (BMI) z-scores are based on a participant's sex and age at each scheduled visit. The World Health Organization (WHO) growth charts will be used to derive z-scores (WHO 2000). If participants become at least 24 months of age, the Center for Disease Control (CDC) growth charts will be used to derive z-scores (CDC 2000).

- **Head Circumference Z-Scores**
  Head circumference z-scores are based on a participant's sex and age at each scheduled visit. The World Health Organization (WHO) growth charts will be used to derive z-scores (WHO 2000).

- **MUAC Z-Scores**
  MUAC z-scores are based on a participant's age at each scheduled visit and can be calculated only from age 2 months on (Abdel-Rahman et al. 2017).

- **Skinfold Thickness Z-Scores**
  Skinfold thickness z-scores are based on a participant's sex and age at each scheduled visit and can be calculated only from age 3 months based on WHO charts

- **Length of Stay for Hospitalizations (days) =** Date of discharge – Date of admission + 1 day

- **Time to Liver-Associated Event (days)** = Date of liver-associated event – Date of first dose of study drug

6.1.9. Data Adjustments/Handling/Conventions

All collected data will be presented in listings. Data not subject to analysis according to this plan will not appear in any tables or graphs but will be included in the data listings.

All p-values will be displayed in four decimals and rounded using standard scientific notation (e.g., 0.xxxx). If a p-value less than 0.0001 occurs, it will be shown in tables as <0.0001.

Participant Age

The age of a participant at screening will be used to determine the age-specific module or instrument for the appropriate assessment (i.e., [redacted]). The same module or instrument will be completed for the duration of the study, regardless of subsequent birthdays throughout the study.

Adverse Event and Concomitant Medication Coding

AEs will be coded using the Medical Dictionary for Regulatory Activities using (MedDRA) version 26.1. Prior and concomitant medications will be coded using World Health Organization Drug Dictionary (WHO-DD Enhanced version September 2019), Anatomical Therapeutic Chemical (ATC) level 2 for ATC class and level 5 (clinical substance) for preferred term.

Prior and Concomitant Treatment Definition and Data Handling

A concomitant treatment refers to all treatment, including concomitant therapies as well as herbal treatments, vitamins, behavioral treatment, and nonpharmacological treatment such as psychotherapy taken between the dates of the first dose of study drug and the end of the participant's participation in the study, inclusive. For participants not continuing into the OLE,

this period of participation is from the first day of screening through the last contact. For participants continuing into OLE, this period is from the first day of screening through the Week 104 visit.

Treatments that started before the first dose of study drug are considered prior treatments whether or not they were stopped prior to the first dose of study drug. Any treatment continuing or starting after the first dose of study drug will be considered as concomitant. If a treatment starts prior to the first dose of study drug and continues after the first dose of study drug, the medication will be considered as both prior and concomitant.

TEAE Data Handling

If an event worsens in severity during the study, the lower grade event is marked as "Not recovered/not resolved" on the AE case report form (CRF) and an end date entered. A new event is recorded on the AE CRF with a start date that matches the end date, and the grading increases (e.g., from Grade 2 to Grade 3). If an event becomes serious, the date that the event became serious is recorded on the AE CRF as the End Date of that AE and the Start Date of the corresponding serious adverse event (SAE).

Adverse Events of Special Interest

An AESI is one of scientific and medical interest specific to the sponsor's product or program.

The following events are AESIs for participants in this study and must be reported to the sponsor within 24 hours after awareness, irrespective of regulatory seriousness criteria or causality:

- LSV deficiency requiring study drug discontinuation (per Section 7.4.5 in the Protocol)
- Liver parameter disruption requiring study drug interruption and/or dose modification (per Section 7.4.2.3 in the Protocol)
- Any events that are suspected and/or confirmed to be due to PG toxicity, which includes but is not limited to neurological complications, hemolysis, and cardiac arrhythmias

Partial Date Imputation

If partial dates are reported, the convention for replacing missing dates for the purpose of calculating derived variables is as follows:

Partial BA Diagnosis Dates

For partial original BA diagnosis dates: a) if only the day is missing, and the month and year match the first dose date, then the day is assigned the first day of the month (01); otherwise the day assigned is 15; and b) if both the day and month are missing, then the day/month assigned is the first day of July (01JUL), as long as the date is before the first dose date; otherwise, the day/month assigned is the first day of January (01JAN). If any of the imputed dates are prior to the birth date, then the day of diagnosis is imputed as the birthday.

Partial AE or Medication Dates

AEs or medications with entirely missing start dates will be classified as treatment emergent or concomitant, as appropriate.

For partial AE or concomitant treatment start dates: a) if only the day is missing and the month and year match the first dose date and the end date is on or after the first dose date, then the date is assigned the first dose date; thus, the event/medication will be considered as treatment emergent/concomitant; if the month and/or year do not match the first dose date or the end date is prior to the first dose date, then the day is assigned the first day of the month (01); b) if both the day and month are missing and the year matches the first dose date and the end date is on or after the first dose date, then the date is assigned the first dose date; if the year does not match the first dose date or the end date is prior to the first dose date, then the day/month are assigned the first day of the year (01JAN).

For partial end dates: a) if only the day is missing, then the day is assigned the last day of the month; b) if both day and month are missing, they are assigned the last day of the year (31DEC).

Partial Liver-Associated Event Date

For partial liver-associated event date: a) if only the day is missing, then the day is assigned the first day of the month (01); b) if both the day and month are missing and the year matches the first dose date, then the date is assigned the first dose date; if the year is after the first dose date, then the day/month are assigned the first day of the year (01JAN).

Dates of Birth

Partial or complete dates of birth are not reported by the investigative sites. Complete date of birth is required, however, to derive a participant's weight, height, and BMI z-scores. The convention for imputing missing birth dates for the purpose of statistical analysis will be based on a participant's age in days at baseline, as defined in Section 6.1.8.

Lower and Upper Limit of Quantitation

In general, for quantitative laboratory values reported as "<" or "≤" the lower limit of quantitation (LLOQ), one-half of the reported value (i.e., LLOQ/2) will be used for analysis. The exception to this data treatment is for plasma maralixibat concentrations that are reported as <LLOQ, where a value of zero will be used in calculating summary statistics.

For quantitative laboratory values reported as ">" or "≥" the upper limit of quantitation (ULOQ), the reported value (i.e., ULOQ) will be used for analysis.

Laboratory Test Results

For analysis purposes, repeat laboratory test results will not be used unless the original laboratory value is missing or indicated as invalid, in which case the first nonmissing repeat laboratory value will be used for data analysis.

Local laboratory results are used where central laboratory results are not available.

The International System of Units (SI) will be used in reporting all efficacy and safety laboratory values, unless otherwise specified in Appendix 1.

Dose Used in Safety Analysis

For all safety and tolerability analyses, participants will be analyzed by the treatment received. For the maralixibat treatment group, the dose of IP received at the end of the dose-escalation period or the last dose received if the participant discontinued from the IP during the dose escalation period will be used for the safety analysis, where specified.

Treatment Received in Adverse Event Listings

For all AE listings, treatment received at the start of the event will be presented. For participants on maralixibat, dose in units of µg/kg/day will be reported. For AEs that started within 7 days after the last dose of study drug, the last treatment received is listed. For AEs that started >7 days after the last dose of study drug, treatment received is blank. For AEs that started prior to the first dose of study drug, NT (Not Treated) will be listed.

Treatment Duration and Exposure

For participants who are missing the date of last study drug application, for any reason, the last known contact date will be used in the calculation of treatment duration and study drug exposure. Study drug compliance will not be calculated for those participants whose date of last study drug administration is unknown.

Censoring for Time to First Liver-Associated Event

If a participant does not have a liver-associated event reported, then the time to first liver-associated event is censored at the last contact date.

7. STUDY PARTICIPANTS AND DEMOGRAPHICS

Participant disposition, demographics, disease history and baseline disease characteristics, important protocol deviations, prior medications, and study drug exposure and compliance will be summarized, unless otherwise noted, by randomized treatment group and overall in the Safety Population.

7.1. Disposition of Participants and Withdrawals

Participant disposition will be tabulated in all participants, including screen failures.

Participant disposition will include tabulations of the number and proportion of participants in each of the analysis populations, completed study treatment, and discontinued early from the study (along with reasons for withdrawal). Percentages will be based on the number of participants in the Safety Population. The participant disposition tabulation will also include the

number of participants screened for eligibility, the number of screen failures, and the number of participants randomly assigned to study treatment.

The number and proportion of randomized and completed participants by country and investigative site, individually, will also be tabulated. Percentages will be based on the number of participants randomly assigned to study treatment.

Study drug accountability and compliance listings will be prepared for all participants in the safety population, showing when the planned dosing schedule was not followed, along with the date and type of dosing deviation. Other disposition and study conduct information, including important protocol deviations, will be listed.

7.2. Protocol Violations and Deviations

Protocol deviations will be tracked, recorded, and reviewed prior to database lock, following the Protocol Deviation Guidance Plan (PDGP) for this study.

Protocol deviations will be classified as "Important" or "Not Important." An important deviation poses as a relevant operational/study conduction deviation, a possible safety issue to the participant, or it has a potential impact on the statistical analysis of the clinical data. A nonimportant deviation is identified as any protocol deviation that does not meet the criteria for an important deviation. Potentially important deviations will be reviewed to determine the final classification as per the final PDGP.

The number and proportion of participants with important protocol violations/deviations will be tabulated by category/type and treatment group in the Safety Population for all participants. These protocol violations/deviations will also be presented in a participant listing by randomized treatment group.

The final decision regarding inclusion and exclusion of participants from the analysis populations will be based on a final listing of protocol deviations. This will be determined during a (blinded) review meeting before any unblinding occurs or database freeze/lock, with input from the Clinical and Biostatistics team members.

In addition, inclusion and exclusion criteria not met and reasons for screen failures will be listed.

7.3. Demographics and Baseline Characteristics

Summary statistics for age at baseline, sex, race, ethnicity, region, height, height z-score, weight, weight z-score, BMI, and BMI z-score will be presented.

Tabulations for age group categories, defined as <median ≥median months of age, will also be presented. Unless otherwise noted, age is the participant's age at the baseline visit for all evaluations and presentations.

Participants reporting >1 race will be counted in a "More than one race" category for purposes of tabulating summary statistics for race.

Disease History and Baseline Disease Characteristics

Summary statistics will also be presented for the following baseline variables:

- Time since original diagnosis of BA (days)
- Participants with baseline UDCA usage
- Participants above and below the median baseline TSB level
- Participants above and below the median baseline sBA level
- Baseline levels of biochemical markers of cholestasis and liver disease and other important baseline laboratory tests:
  - Total sBAs
  - ALT
  - ALP
  - AST
  - GGT
  - Platelets
  - Bilirubin (total and direct)
  - [redacted]
  - 25-hydroxyvitamin D
  - Alpha tocopherol
  - Prothrombin intl. normalized ratio
  - Vitamin A
  - APRI
  - PELD score

Participant demographics and baseline characteristics and medical and surgical history information and prior medications will be presented in a participant listing by randomized treatment group.

7.4. Prior and Concomitant Medications/Therapies

Prior and concomitant treatments will be summarized descriptively by treatment group through use of the number and proportion of participants by ATC Class Level 2 (therapeutic main group) and ATC Class Level 5 (chemical substance).

Prior treatments will be presented separate from concomitant treatments.

7.5. Exposure and Compliance

Treatment exposure will be calculated for each participant exposed to maralixibat during the study and will be summarized descriptively. This analysis will be conducted for the overall 26-week treatment period and for the LTE period, separately.

Treatment exposure summaries will include total treatment duration (days), treatment exposure (days), total dose (μg/kg), and average daily dose (μg/kg/day).

Treatment compliance (%) will also be summarized. For a given day, a participant is considered compliant with treatment if any amount of study drug was administered. In addition to the presentation of descriptive statistics on compliance rates, treatment compliance will also be categorized and summarized as <80%, 80%–<90%, and 90%–100%.

Study drug accountability will be presented in a participant listing by randomized treatment group. Dosing details will be listed separately.

8. EFFICACY ANALYSIS

Efficacy analyses will be conducted in the ITT population. Analysis of the primary efficacy endpoint and secondary efficacy endpoints that achieve statistical significance in the hierarchical testing will also be performed in the PP population.

Analyses on the exploratory efficacy endpoints will be performed in the ITT population.

For all efficacy analyses, participants will be analyzed by the randomized treatment group assignment (maralixibat or placebo), based on the ITT principle that asserts that the effect of a treatment can be best assessed by evaluating on the basis of the intention to treat a participant (i.e., the planned treatment regimen) rather than the actual treatment given. All efficacy data will be presented in participant listings.

For all analyses, site will also be used as a covariate in the MMRM analysis.

8.1. Primary Efficacy Analysis

For this study, the primary estimand is the improvement in TSB measured as change from baseline in the average TSB in the maralixibat treatment group relative to the placebo group. In the course of the 26-week randomized treatment period, participants may be exposed to possible known or unknown intercurrent events that could possibly impact the interpretation of the measures associated with the clinical question of interest, such as treatment discontinuation due to a specific adverse effect or perhaps a lack of effect. The “Hypothetical Strategy” has been adopted for handling all known intercurrent events in this study. To this end, a restricted maximum likelihood (REML)-based MMRM model conducted in the ITT population will be used as the primary analysis method.

The repeated measures include postbaseline time periods during the dose escalation period (i.e., Weeks 1–8) and stable dosing period (i.e., Weeks 9–26), with change from baseline in TSB as the dependent variable. The MMRM model will include the fixed, categorical effects of

treatment group, site, time period, and treatment group–by-time period interaction as well as the continuous, fixed covariates of baseline TSB and the baseline score-by-time period interaction. The MMRM analysis method is further described in Section 6.1.4.1.

The unstructured variance/covariance matrix will be used to model the variances and covariances for the six time points included in the model. The unstructured variance/covariance does not impose any restrictions on the pattern of the matrix elements. If there is a convergence issue with the unstructured covariance model, the following variance/covariance matrix structures will be used in the following order: 1) heterogeneous Toeplitz, 2) heterogeneous autoregressive of order 1, and 3) heterogeneous compound symmetry. The first variance/covariance structure that does not have a convergence problem will be the one used for the analysis. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom (Kenward and Roger 1997).

The primary efficacy analysis will compare maralixibat and placebo with use of the contrast (difference in least squares [LS] means) between treatment groups across the last 12 weeks of the study [redacted] The analytical solution of the overall treatment effect obtained from MMRM is an equally weighted average of the 3 individual visit-specific estimates over the time period of interest [redacted] Significance tests will be based on LS means using a 2-sided significance level (2-sided 95% CIs).

The null hypothesis for the primary efficacy endpoint of the equality of maralixibat and placebo is:

> H_{01}: Mean changes in TSB between baseline and [redacted] in the two treatment groups are equal

The null hypothesis of equal treatment effect will be rejected if the statistical analysis results in a 2-sided p-value for treatment over the [redacted] of < 0.05. LS means will be calculated for each treatment group for each postbaseline time period in the model. The difference between maralixibat and placebo change from baseline in TSB will be estimated, with the corresponding 2-sided 95% CI constructed for each time period and [redacted] of the study combined. The change from baseline LS means with standard error, 95% CI for the LS means, p-value for testing if the LS mean is zero, LS mean difference between treatment groups (maralixibat minus placebo) with standard error, 95% CI for the LS mean difference, and p-value for testing if the treatment LS means are equal will be presented.

The trial will be claimed successful if the hypothesis of no treatment effect on the primary efficacy endpoint in the ITT population is rejected at the 0.05 (2-sided) significance level.

8.2. Secondary Efficacy Analysis

Analyses similar to that described for the primary efficacy endpoint will be performed for each of the change from baseline secondary efficacy endpoints. For categorical endpoints, the number and proportion of participants will be summarized by treatment group for each analysis visit or time period, as appropriate. Barnard's exact unconditional test will be used to calculate the p-value for the difference between treatment groups.

The null hypotheses for the secondary efficacy endpoints of the equality of maralixibat and placebo are:

H_{02}: Mean changes in total sBA level between baseline and Week 14 through Week 26 in the two treatment groups are equal

H_{03}: Proportions of participants with mean TSB levels <2 mg/dL at Weeks 14, 20, and 26 in the two treatment groups are equal

H_{04}: Proportions of participants observed to have a liver-related clinical event, including liver transplantation, liver decompensation (hepatic encephalopathy, variceal bleeding, new persistent ascites), discontinuations due to liver related events, or death through Week 26 in the two treatment groups are equal

H_{05}: Proportions of participants undergoing liver transplantation or death through Week 26 in the two treatment groups are equal

H_{06}: Proportions of participants observed to develop clinically evident portal hypertension defined as splenomegaly (spleen size >2 cm below the costal margin palpated on physical examination) and thrombocytopenia (platelet count <150 x 10^9/L) or clinically evident ascites or endoscopic evidence of esophageal or gastric varices through Week 26 in the two treatment groups are equal

H_{07}: Proportions of participants with mean TSB levels ≤1.2 mg/dL at Weeks 14, 20, and 26 in the two treatment groups are equal

H_{08}: Proportions of participants with mean sBA levels <40 mmol/L at Weeks 14, 20, and 26 in the two treatment groups are equal

A hierarchical testing procedure, as described in Section 6.1.3, will be used in the comparisons between maralixibat and placebo on the primary and secondary efficacy endpoints in the ITT population.

All tests will be performed as 2-sided tests at the 0.05 level of significance.

8.3. Sensitivity and Supportive Analyses for Primary and Secondary Endpoints

Sensitivity and/or supportive analysis will be performed on the primary and selected secondary efficacy endpoints to quantify the possible impact of missing data and to demonstrate the robustness of the conclusions.

Sensitivity analysis will be performed on the efficacy endpoints in the ITT analysis population that achieve statistical significance in the hierarchical testing. These endpoints will be also evaluated in the PP analysis population.

The following sensitivity analyses will be performed on the primary and first two secondary analyses only if they achieve statistical significance. Although the assumption of MAR, as used for the primary efficacy analysis method, is often reasonable in clinical trials, the possibility of MNAR data cannot be ruled out. Sensitivity analyses to deal with missing data will use MI methods where missing values are imputed individually under both a plausible MAR and MNAR scenario. Two sensitivity analysis models, one based on the standard MAR imputation approach and the other on a tipping-point analysis, will be used to examine robustness of the primary analysis results (see Section 6.1.4.2).

Summary statistics on the primary efficacy endpoint and the second secondary efficacy endpoint will also be presented descriptively by treatment group and visit, overall and for each of the following subgroups in the ITT analysis population:

- Region (Asia, Europe, Middle East, North America, and South and Central America)
- Age at baseline (< median, ≥ median)
- Sex (male, female)
- Race (White, Asian, Black or African American, American Indian or Alaska Native, multiple)
- Baseline TSB value (< median, ≥ median)
- Baseline serum bile acid (< median, ≥ median)

8.4. Adjustment for Multiplicity

A hierarchical testing procedure will be used in the comparisons between maralixibat and placebo on the primary and secondary efficacy endpoints in the ITT population using the primary analysis method (as described for the primary endpoint). The hierarchical order for testing the null hypotheses is listed in Section 6.1.3.

The effect of such a procedure maintains study-wide Type I error control, ensuring that no confirmatory claims can be based on the endpoint(s) that have a testing rank lower than that endpoint whose null hypothesis was the first that could not be rejected. Those tests will be considered as exploratory without any type I error control for the additional subsequent tests.

8.5. Exploratory Efficacy Analysis

Analyses similar to that described for the primary efficacy endpoint (i.e., MMRM) will be performed for each exploratory endpoint that is based on change from baseline values on continuous variables. LS means will be calculated for each treatment group for each postbaseline visit in the model. The difference between maralixibat and placebo change from baseline in each outcome variable will be estimated, with the corresponding 2-sided 95% CI constructed for each visit. Change from baseline LS means with standard error, 95% CI for the LS means, p-value for testing if the LS mean is zero, LS mean difference between treatment groups (maralixibat minus placebo) with standard error, 95% CI for the LS mean difference, and p-value for testing if the treatment LS means are equal will be calculated for each visit. For the endpoints that compare treatment groups across the last 12 weeks of the study, the same methods as described for the primary analysis of the primary efficacy endpoint will be applied. The number of observations, mean, 95% CI on the mean, SD, median, minimum, and maximum on

both the observed and change from baseline values will also be summarized by treatment group for each visit, including the end of study visit (Week 26/EOT).

For responder-type endpoints (i.e., efficacy variables involving binary outcomes), the number and proportion of participants who are considered a "responder" will be summarized by treatment group for each analysis visit or time period, as appropriate. Barnard's exact unconditional test will be used to calculate the p-value for the difference between treatment groups.

For responder-type endpoints with ordinal measures, the Cochran-Mantel-Haenszel (CMH) test will be applied to test for no association between treatment group and the variable of interest.

The number and proportion of participants with total sBA % decrease from baseline to Weeks 2, 6, 14, 20, and 26 within the following categories: >75% (large decrease), 0%–75% (decrease), and <0% (increase) will be analyzed using a CMH test to test for no association between treatment group and the variable of interest at each analysis visit.

Time to first liver-associated event (listing for liver transplantation, development of liver cirrhosis, development of portal hypertension, spontaneous bacterial peritonitis, hepatocellular carcinoma, cholangiocarcinoma, liver decompensation [hepatic encephalopathy, variceal bleeding, new persistent ascites], and death) will be compared among treatments through use of Kaplan-Meier product-limit survival curve estimates. The median and other quartiles for time to first liver-associated event, along with 2-sided 95% CIs, will be estimated based on the Kaplan-Meier method. Both a log-rank test and a Wilcoxon test will be used to test for treatment differences.

Primary, secondary, and exploratory endpoints will be performed using the TIP, if possible.

All tests will be performed as 2-sided tests at the 0.05 level of significance.

Sensitivity analyses are not planned to be performed on the exploratory efficacy endpoints.

9. SAFETY AND TOLERABILITY ANALYSIS

All safety analyses will be performed on the Safety Population, defined as all participants who were randomized and received at least 1 dose of study drug.

Safety measures including AEs, clinical laboratory values, physical examination findings (including body weight, height, and BMI), vital signs, ECGs, and concomitant treatment usage will be summarized descriptively. No inferential statistical tests will be performed, unless otherwise specified. For quantitative variables, descriptive statistics including number of observations, mean, median, Q1, Q3, SD, minimum, and maximum will be presented for observed and change from baseline values at each study visit. Qualitative variables will be summarized using counts and percentages.

Safety data collected at the baseline visit (VBL/Day 0) or the last preceding visit if not collected at VBL will be used as the baseline value for safety analyses.

All safety and tolerability data will be presented in participant listings.

9.1. Adverse Events

All summaries of AEs will be based on TEAEs unless specified otherwise. TEAEs are defined as described in Section 6.1.8 (Derived Variables).

AEs will be coded using MedDRA. The number and proportion of participants who experience the event according to MedDRA system organ class (SOC) and preferred term (PT) will be presented by treatment group. TEAEs related to study drug, AEs that led to withdrawal, SAEs, deaths, and AESIs will be similarly summarized.

A summary of TEAEs will be presented by treatment group. The summary will include the total number and percent of participants reporting:

- Any TEAEs
- Any treatment-related TEAE
- Any severe TEAE
- Any severe treatment-related TEAE
- Any serious TEAE
- Any serious treatment-related TEAE
- Any TEAE that led to permanent study drug discontinuation
- TEAEs that resulte in death
- Any AESI

A participant with multiple reported cases of the same AE will be counted once within each SOC and similarly counted once within each PT. TEAEs summarized by SOC and PT will be sorted in alphabetical order of the SOC and by descending frequency order of the PT within each SOC.

In addition, the following subgroup analyses will be explored for the summary of TEAEs and the incidence of TEAEs by SOC and PT indicated above:

- Age at time of Kasai (<median and ≥median)
- Sex (male, female)
- Race (Asian and non-Asian)

Missing and partially missing AE start and/or stop dates will be imputed, for the purpose of statistical analysis, according to the specifications described in Section 6.1.9.

Listings of AEs will be presented in by-participant listings, detailing the treatment received at the start of the event, including dose for active study drug, SOC, PT, verbatim term given by the investigator, onset date and study day, end date and study day, event duration, severity, relationship to study drug, outcome, action taken with study drug, seriousness, and treatment required. Events that are treatment emergent will be flagged.

9.1.1. Adverse Events That Led to Discontinuation of Study Drug

AEs that led to permanent discontinuation of study drug will be tabulated by treatment group. Participant listings of AEs that led to permanent discontinuation of study drug will also be presented.

9.1.2. Deaths and Serious Adverse Events

Treatment-emergent SAEs will be summarized in the same manner as AEs that led to permanent discontinuation of study drug. Participant listings of all SAEs will also be presented.

Any deaths that occur during the study will be presented in a participant listing. The listing will include participant ID, study drug and dose received at the time of death (or the last study drug/dose received prior to death), date of death, number of days between the first and last dose, MedDRA PT, and relationship to study drug.

9.2. Clinical Laboratory Evaluations

Safety laboratory test results will be summarized descriptively by study visit and treatment group as observed and change from baseline values. The number and proportion of participants with clinical laboratory values below, within, or above the normal range by time point and in relation to baseline will be tabulated for each select lipid-soluble vitamin (LSV) laboratory analytes by treatment group in a shift table. The shift table will include the following LSVs: 25-hydroxyvitamin D (ng/mL), alpha tocopherol (mg/dL), INR, and vitamin A (µg/dL).

All safety laboratory test parameters will be presented by panel in participant listings.

Efficacy laboratory tests will not be included in safety summaries or listings. A separate participant listing for efficacy laboratory results will be presented.

9.3. Vital Signs and Weight/Height Measurements

Vital signs (temperature, systolic and diastolic blood pressure, heart rate, and respiratory rate), weight, height, and BMI will be summarized descriptively by study visit and treatment group as observed and change from baseline values. Weight, height, and BMI measurements will also be summarized as a z-score for a participant's age and sex.

9.4. Electrocardiograms

Descriptive summaries will be presented for ECG measures of PR interval, QRS duration, and QT interval corrected using both Bazett's and Fridericia's formula (QTcB and QTcF). The number and proportion of participants with normal, abnormal-not clinically significant, and abnormal-clinically significant ECG results will also be summarized. These summaries will be presented by study visit and treatment group.

9.5. Concomitant Medications/Therapies

Concomitant treatments will be summarized descriptively by treatment group with use of the number and proportion of participants by ATC Class Level 2 (therapeutic main group) and ATC Class Level 5 (chemical substance).

Prior therapies will be presented separate from concomitant therapies. Any treatments continuing or starting after the first dose of study drug will be considered concomitant. If a therapy starts prior to the first dose of study drug and continues after the first dose of study drug, it will be considered both prior and concomitant.

10. OTHER PLANNED ANALYSIS

Healthcare utilization and PK analyses will be performed on the Safety Population.

10.1. Healthcare Utilization

The number of outpatient clinic visits and hospitalizations related to underlying disease, length of stay for hospitalizations, Emergency Department visits (days), and any surgeries and procedures specific to the participant's condition will be collected during postbaseline visits as outlined in the Table 1 and Table 2. The date of visit/admission, date of discharge, relationship/issue, and discharge status will be collected. The number of days the caregiver missed work due to the events collected for healthcare utilization will also be collected.

Participants that visit the Emergency Department and the visit results in the participant being admitted to the hospital will be counted in both categories.

Number and percent of participants and number of occurrences will be presented across the entire postbaseline study period, between clinic visits from baseline through follow-up, by treatment group for BA-related hospitalizations and ER visits. For length of hospital stays and number of days a caregiver missed work, summary statistics on a participant-level basis will be presented.

10.2. Pharmacokinetic Analyses

Systemic concentrations of maralixibat in plasma will be determined at predose and at approximately 2.5 hours after the morning dose at Week 10 and Week 26 (ET). Summary statistics (number of observations, mean, SD, coefficient of variation, median, minimum, maximum, and geometric mean) will be determined for maralixibat concentrations at each analysis visit.

11. Changes from Protocol Planned Analysis

The order of the secondary endpoints was updated based on further evaluation of their importance.

A new exploratory population, TIP, was defined.

12. REFERENCES

Abdel-Rahman SM, Bi C, Thaete K. Construction of lambda, mu, sigma values for determining mid-upper arm circumference z scores in U.S. children aged 2 months through 18 years. Nutr Clin Pract 2017;32:68–76.

ASA. Ethical guidelines for statistical practice. Prepared by the committee on professional ethics, April 2018. http://www.amstat.org/about/ethicalguidelines.cfm.

Bezerra JA, Spino C, Magee J, et al. Use of corticosteroids after hepatoportoenterostomy for bile drainage in infants with biliary atresia: the START randomized clinical trial. JAMA 2014;311:1850–9.

CDC. Centers for Disease Control growth charts – a SAS program for the 2000 CDC Growth Charts (ages 0 to <20 years). 2000; available at: www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm.

Cnaan A, Laird NM, Slasor P. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. Stat Med 1997;16:2349–80.

European Medicines Agency. Guideline on missing data in confirmatory clinical trials. 2010.

ICH. ICH Harmonised tripartite guideline. statistical principles for clinical trials e9. 1998. https://database.ich.org/sites/default/files/E9_Guideline.pdf.

Kenward M G, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics 1997;53:983–97.

Little R, Rubin D. Statistical analysis with missing data. John Wiley: 1987.

Little R, Yau L. Intent-to-treat analysis for longitudinal studies with drop-outs. Biometrics 1996;52:1324–33.

Mallinckrodt CH, Clark WS, David SR. Accounting for dropout bias using mixed-effects models. J Biopharm Stat 2001;11:9–21.

Mallinckrodt CH, Lane PW, Schnell D, et al. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. Drug Inf J 2008;42:303–19.

Mallinckrodt CH, Roger J, Chuang-Stein C, et al. Recent developments in the prevention and treatment of missing data. Drug Inf J 2013;48:68–80.

Molenberghs G, Kenward MG. Missing data in clinical studies. Wiley: 2007.

Molenberghs G, Thijs H, Jansen I, et al. Analyzing incomplete longitudinal clinical trial data. Biostatistics 2004;5:445–64.

PedsQL (2021). The PedsQL measurement model for the pediatric quality of life inventory – scoring instructions; available at: www.pedsql.org.

Royal Statistical Society. Code of conduct. 2014; available at: http://www.rss.org.uk/Images/PDF/join-us/RSS-Code-of-Conduct-2014.pdf.

Schafer JL. Analysis of incomplete multivariate data. Chapman & Hall 1997.

Siddiqui O, Hung HM, O'Neill R. MMRM vs. LOCF: a comprehensive comparison based on simulation study and 25 NDA datasets. J Biopharm Stat 2009;19:227–46.

Verbeke G, Molenberghs G. Linear mixed models for longitudinal data. Springer 2000.

WHO. World Health Organization growth charts – A SAS Program for the WHO growth charts (ages 0 to <2 years). 2000; available at www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas-who.htm.

13. TABLES, LISTINGS, AND FIGURES

All listings, tables, and figures will have a header showing the company name (Mirum Pharmaceuticals, Inc.) and protocol (MRX-701), a page footer showing the file name and path, and the date/time of program execution along with the database data cut date.

The following reporting conventions will be adopted for the presentation of study data. These conventions will enhance the review process and help to standardize the presentation with common notations.

General Reporting Conventions

- All tables and data listings will be developed in landscape orientation.
- Specialized text styles, such as bolding, italics, borders, shading, and superscripted and subscripted text, will not be used in tables, figures, and data listings unless they add significant value to the table, figure, or data listing.
- Only standard keyboard characters should be used in tables and data listings. Special characters, such as nonprintable control characters, printer-specific characters, or font specific characters, will not be used on a table, figure, or data listing.
- AEs with missing MedDRA coding will have their System Organ Class and/or Preferred Term presented as "Not Coded" in the tables. The "Not Coded" frequencies will be sorted to the end of the tables. AEs that are not coded are not expected.
- Programming notes may be inserted into the shells; these notes will not appear in the final output.

Population Summary Conventions

- Population sizes may be presented for each classification factor as totals in the column header as (N=xx), where appropriate.
- All population summaries for categorical variables will include all categories that were planned and for which the participants may have had a response. Percentages corresponding to null categories (cells) will be suppressed; however, counts and percentages of missing values may be needed. Counts of zero will be presented as "0"—without the percentage.
- All population summaries for continuous variables will include n, mean, SD, median, Q1, Q3, minimum, and maximum. Other summaries (e.g., number missing, geometric mean, 95% CIs, and coefficient of variation [CV] or % CV) may be used as appropriate. The precision of the maximum and minimum will match the maximum precision in the data. The mean and median will have 1 additional decimal place unless specified. The SD will have 2 additional decimal places.
- All percentages are rounded and reported to a single decimal point (xx.x%).

13.1. Planned Table Descriptions

The following are planned summary tables for Study MRX-701. The table numbers and titles are place holders only and will be determined when the tables are produced. The efficacy outputs are ordered to account for the hierarchical testing structure as outlined in Section 6.1.3,

where possible. For cases where multiple study visits for the same parameter are tested, those visits will be consolidated on a single table for ease of review and referencing results. Similarly, for cases where multiple analysis populations or cohort groups are presented, this data will be consolidated on a single table, when practical.

The table shells will be provided under a separate document.

13.1.1. Participant Data

Table 5 Participant Data Summary Tables

Number	Population	Title
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

13.1.2. Efficacy Data

Table 6 Efficacy Data Summary Tables

Number	Population	Title
[redacted]		
[redacted]	[redacted]	[redacted]
[redacted]		
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

Number	Population	Title
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]		
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

Number	Population	Title
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
14.2.4.x Exploratory Analyses		
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

Number	Population	Title
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

Number	Population	Title
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

13.1.3. Safety Data

Table 7 Safety Data Summary Tables

Number	Population	Title
[redacted]		
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]		
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]		
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]		
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

Number	Population	Title
[redacted]	[redacted]	[redacted]
[redacted]		
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]		
[redacted]	[redacted]	[redacted]

13.1.4. Other Data

Table 8 Other Data Summary Tables

Number	Population	Title
[redacted]		
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

Additional tables may be added post-hoc to further examine study data.

13.2. Planned Figure Descriptions

The following are planned summary figures for Study MRX-701. The figure numbers are place holders only and will be determined when the figures are produced.

Estimates from the MMRM model (over time) will be graphically displayed for select efficacy variables in the ITT population. These plots will display the LS Mean ±SE for each treatment group at each analysis visit/time period, as appropriate. Each of the treatment group LS means will be displayed side by side, with vertical lines emanating from each LS Mean to represent the SE. At minimum, these graphical displays will be presented for the change from baseline in total bilirubin and change from baseline in sBA levels, over time. The number of participants with nonmissing change from baseline values, for each treatment group and analysis visit, will be displayed along the horizontal axis at each time point, baseline through Week 26.

Tipping-point results, for the sensitivity analysis of the primary efficacy endpoint and selected secondary efficacy endpoint, will be presented in a heatmap. This figure will be generated to display p-values corresponding to all combinations of shifts from the imputed data. For the heatmap figure, each cell will be colored according to the p-value: green for p-values <0.05, yellow for p-values of 0.05 to <0.10, orange for p-values of 0.10 to <0.20, and red for p-values of 0.20 to 1.00.

Additional figures may be added post hoc to further examine study data.

13.2.1. Efficacy Data

Table 9 Efficacy Data Summary Figures

Number	Population	Title
[redacted]		
[redacted]	[redacted]	[redacted]
[redacted]		
[redacted]	[redacted]	[redacted]
[redacted]		
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

13.3. Planned Listing Descriptions

In general, listings produced will include the participant data collected on associated CRF pages. All listings will be sorted by treatment and participant ID. For safety data (i.e., study drug exposure, AEs, vital signs, laboratory tests, ECG, physical examination, and plasma maralixibat

concentration levels), listings will be presented by treatment received. For all other listings (including efficacy laboratory tests), where applicable, data will be presented by randomized treatment group.

For listings of efficacy variables, change from baseline values will be included, as appropriate.

For AE listings, study day relative to the date of first dose of study drug will be provided along with stop and start dates. The dose of study drug at the onset of AEs in the maralixibat treatment group will also be provided. For partial event dates, study day will be derived using an imputed date as described in Section 6.1.9. The partial event data will be presented in the listing.

Study day relative to the first dose of study drug will also be included on laboratory listings, along with laboratory collection dates, and on study drug exposure listings, along with start and stop dates.

For the concomitant treatment listing, medications that are ongoing at the time of informed consent will be indicated.

Weight, height, and BMI will be presented in the vital signs listing, rather than the physical examination listing.

[redacted] listings will present total scale and summary scores, including [redacted]

Assessments denoted as clinic visit "Week 26/ET" in listings will include assessments performed at the Early Termination visit for participants who withdraw early from the study.

In all listings, a blank line will be placed between participants. Within a data listing, if an item appears line after line (e.g., repetition of participant number), then only the first occurrence will be displayed.

The listing shells will be provided under a separate document.

13.4. Standard Layout for all Tables, Listings, and Figures

The following standard layout will be applied to all Tables, Listings, and Figures (TLFs) in support of this study. Note that programming notes may be added if appropriate after each TLF shell.

Figure 2 Standardized Layout

Mirum Pharmaceuticals, Inc. Protocol: MRX-701	CONFIDENTIAL	Page xx of xx Version xxxxx

<Table, Listing, Figure> xx.x.x

<Title of Table Listing or Figure>

<Study Population and if applicable subgroup Description>

Body of Table, Listing or Figure

<Abbreviations: If applicable>

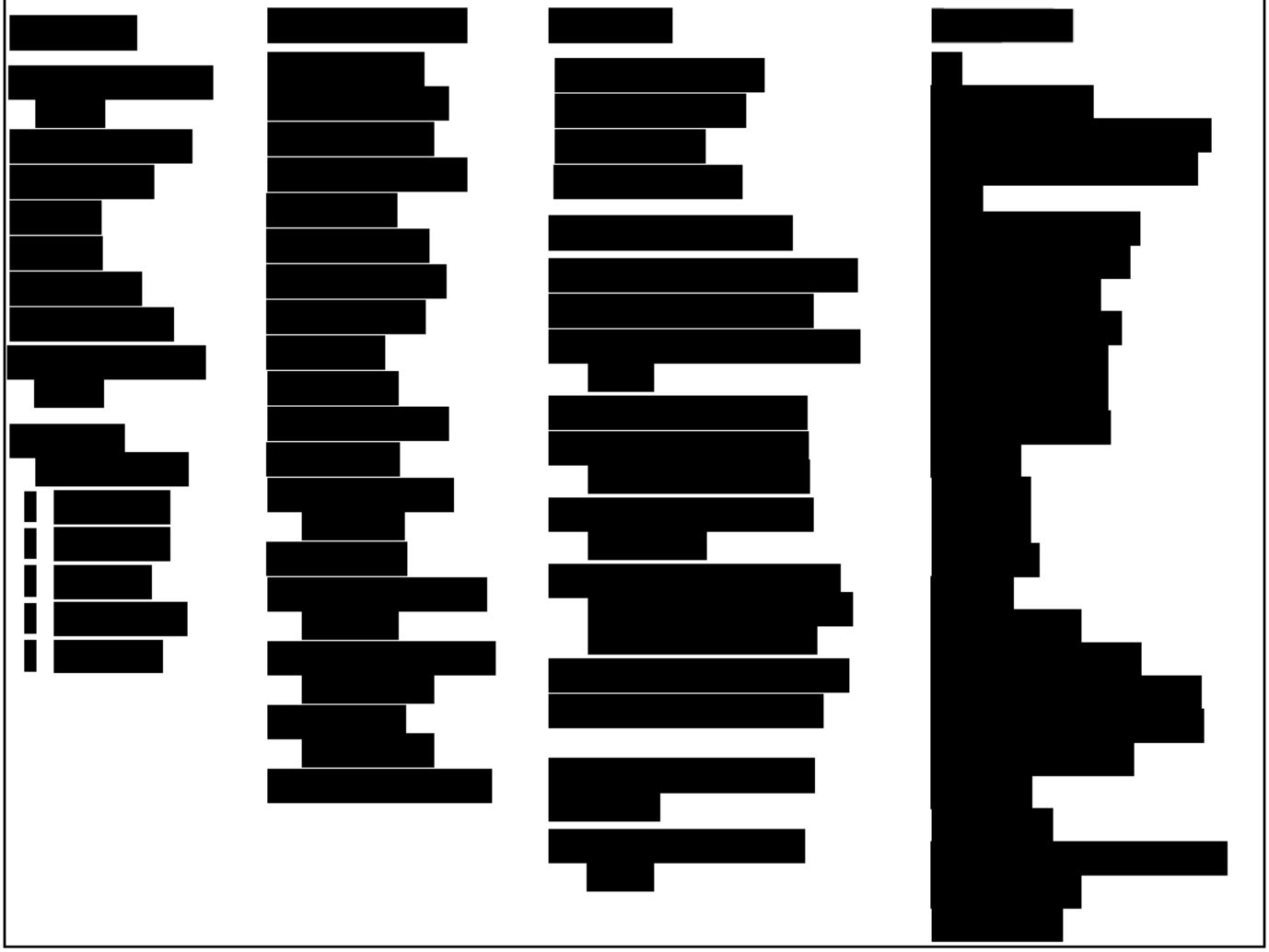
<Note: If applicable>

Footnote 1 <if applicable> Recommendation is to keep footnotes to a minimum
Footnote 2 <if applicable>
Footnote n <if applicable>

<pgm path and name>
<Executed on <date> at <time> on data from <data cut date>

Appendix 1 Listing of Safety and Efficacy Laboratory Analytes

SAFETY LABS



EFFICACY LABS

[1] Listing only
[2] Safety and efficacy lab tests
[3] Calculated lab parameters (see protocol Sec 6.1.8)

Appendix 2 Listing of Lipid-Soluble Vitamin Deficiency Events

The following MedDRA Preferred Terms associated with LSV deficiency events are included as an AESI: