STATISTICAL ANALYSIS PLAN

Study Title:	A Phase 3, Randomized, Double-Blind, Placebo-Controlled Study Evaluating the Safety, Tolerability, and Efficacy of Cilofexor in Non-Cirrhotic Subjects With Primary Sclerosing Cholangitis
Name of Test Drug:	Cilofexor
Study Number:	GS-US-428-4194
Protocol Version (Date):	Amendment 5 (16 March 2022)
Analysis Type:	Interim, Primary, and Final Analysis
Analysis Plan Version:	Version 1.0
Analysis Plan Date:	10 August 2022
Analysis Plan Author(s):	PPD PPD

CONFIDENTIAL AND PROPRIETARY INFORMATION

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS

AE	adverse event
ALP	alkaline phosphatase
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
AOS	Amsterdam-Oxford score
AST	aspartate aminotransferase
BLQ	below the limit of quantitation
BMI	body mass index
C4	7-alpha-hydroxy-4-cholesten-3-one
CI	confidence interval
CILO	cilofexor
CP	Child-Pugh
CRF	case report form
CRP	C-reactive protein
CSR	clinical study report
CTCAE	Common Toxicity Criteria for Adverse Events
DILI	drug-induced liver injury
DMC	data monitoring committee
EAIR	exposure-adjusted incidence rate
ECG	electrocardiogram
eCRF	electronic case report form
EFS	event-free survival
ELF™	enhanced liver fibrosis
ET	early termination
FAS	Full Analysis Set
FGF19	fibroblast growth factor 19
GGT	gamma-glutamyltransferase
HbA1c	hemoglobin A1c
HCC	hepatocellular carcinoma
HDL-C	high density lipoprotein-cholesterol
HLGT	high-level group term
HLT	high-level term
HRUQ	health resource utilization questionnaire
IBD	inflammatory bowel disease
ID	identification
INR	international normalized ratio
IPD	important protocol deviation
IRT	interactive response technology

KM	Kaplan-Meier
LDL-C	low density lipoprotein-cholesterol
LLN	lower limit of normal
LLT	lower-level term
LOCF	last observation carried forward
LOQ	limit of quantitation
LS	least square
MAR	missing at random
MedDRA	Medical Dictionary for Regulatory Activities
MELD	model for end-stage liver disease
MH	Mantel-Haenszel
MI	multiple imputation
MNAR	missing not at random
MRCP	magnetic resonance cholangiopancreatography
MRS	Mayo risk score
NRI	non-responder imputation
OC	observed case
OLE	open-label extention
P3NP	procollagen 3 N-terminal propeptide
PD	pharmacodynamic
PI	principal investigator
PK	pharmacokinetic
PP	predictive power
PRO	patient-reported outcome
PSC	primary sclerosing cholangitis
PT	preferred term
PTM	placebo-to-match
Q1, Q3	first quartile, third quartile
QoL	quality of life
SAE	serious adverse event
SAP	statistical analysis plan
SD	standard deviation
SE	standard error
SOC	system organ class
TE	treatment-emergent
TEAE	treatment-emergent adverse event
TFLs	tables, figures, and listings
TIMP1	tissue inhibitor of metalloproteinase 1
TPA	tipping point analysis
UC	ulcerative colitis

UDCA	ursodeoxycholic acid
ULN	upper limit of normal
VLDL-C	very low density lipoprotein-cholesterol
WHO	World Health Organization

PHARMACOKINETIC ABBREVIATIONS

AUC_{last}	area under the concentration versus time curve from time zero to the last quantifiable concentration
AUC_{tau}	area under the concentration versus time curve over the dosing interval
C_{last}	last observed quantifiable concentration of the drug
C_{max}	maximum observed concentration of drug
C_{tau}	observed drug concentration at the end of the dosing interval
CLss/F	apparent oral clearance after administration of the drug: at steady state: CLss/F = Dose/AUC_{tau} , where "Dose" is the dose of the drug
$t_{1/2}$	estimate of the terminal elimination half-life of the drug, calculated by dividing the natural log of 2 by the terminal elimination rate constant (λ_z)
T_{last}	time (observed time point) of C_{last}
T_{max}	time (observed time point) of C_{max}
λz	terminal elimination rate constant, estimated by linear regression of the terminal elimination phase of the concentration of drug versus time curve

1. INTRODUCTION

This statistical analysis plan (SAP) describes the statistical analysis methods and data presentations to be used in tables, figures, and listings (TFLs) for an interim analysis, as well as the primary and final analyses, for Study GS-US-428-4194. A formal interim analysis for futility will be performed after the first 160 randomized and dosed participants have completed Week 96 or the early termination (ET) visit assessments in the Blinded Study Phase. The primary analysis will be conducted after all participants have completed or prematurely discontinued from the Blinded Study Phase of the study and completed the follow-up visit, if applicable, for the Blinded Study Phase. The final analysis will be performed after all participants have completed the OLE Phase follow-up visit or prematurely discontinued from the study.

This SAP is based on the study protocol amendment 5 dated 16 March 2022 and the electronic case report form (eCRF). The SAP will be finalized prior to data finalization for the interim futility analysis. Any changes made after the finalization of the SAP will be documented in the clinical study report (CSR).

1.1. Study Objectives

The primary objective of this study is as follows:

- To evaluate whether cilofexor (CILO) reduces the risk of fibrosis progression among noncirrhotic participants with primary sclerosing cholangitis (PSC) at Blinded Study Phase Week 96

The secondary objectives of this study are as follows:

- To assess the safety and tolerability of CILO
- To evaluate changes in serum concentrations of alkaline phosphatase (ALP), alanine aminotransferase (ALT), and bile acids at Blinded Study Phase Week 96
- To evaluate whether CILO increases the proportion of participants with $\geq 25\%$ relative reduction in serum ALP concentration from baseline (biochemical response) and no worsening of fibrosis according to the Ludwig classification (histologic response) at Blinded Study Phase Week 96
- To evaluate fibrosis stage improvement at Blinded Study Phase Week 96
- To evaluate changes in noninvasive markers of fibrosis, including liver stiffness by FibroScan and enhanced liver fibrosis test (ELF test) score at Blinded Study Phase Week 96
- To evaluate change in PSC Symptoms – Module 1 based on the disease-specific PSC patient-reported outcome (PSC-PRO) at Blinded Study Phase Week 96

The exploratory objectives of this study are as follows:

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

1.2. Study Design

This is a Phase 3, randomized, double-blind, placebo-controlled study evaluating the safety and efficacy of CILO in participants with PSC without cirrhosis. The study will consist of 2 phases: a Blinded Study Phase and an Open-label Extension (OLE) Phase.

Blinded Study Phase: Includes a 10-week screening period, 96 weeks of treatment, and a Blinded Study Phase follow-up visit 4 weeks after completion of Blinded Study Phase Week 96 or ET visit. Participants meeting the study's entry criteria will be randomly assigned in a 2:1 ratio to receive CILO 100 mg orally once daily or placebo. Central randomization is used. Randomization will be stratified by the presence or absence of ursodeoxycholic acid (UDCA) use and presence or absence of bridging fibrosis (Ludwig fibrosis score, F3 versus F0-F2) on the screening liver biopsy. Study drug will be administered for a total of 96 weeks from the baseline/Day 1 visit.

Participants who do not permanently discontinue study drug and who complete the Blinded Study Phase Week 96 with an evaluable biopsy (noncirrhotic F0-F3) as determined by the central reader and Blinded Study Phase follow-up visit will be eligible to enter into the OLE Phase.

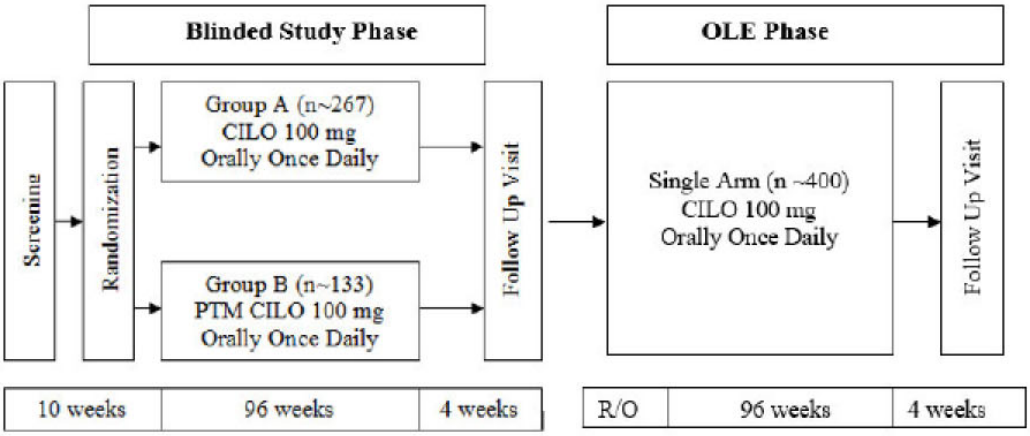
Interim Futility Analysis:

There will be 1 planned interim futility analysis based on the primary endpoint after the first 160 randomized and dosed participants have completed Week 96 or ET assessments in the Blinded Study Phase. A predictive power (PP) approach will be used for futility assessment. The DMC may recommend early termination of the study due to futility if the criterion of PP $\leq$ 10% is met.

If the DMC recommends early termination of the study, a restricted group of senior management at Gilead will be unblinded to evaluate the study data per the interim analysis communication plan. Unblinding of specific Gilead personnel will be documented per the appropriate standard operating procedures (SOPs).

Open-label Extension Phase: Includes 96 weeks of open-label treatment of CILO 100 mg orally once daily and an OLE Phase follow-up visit 4 weeks after completion of the OLE Phase Week 96 visit or ET visit.

The overall study design is presented graphically in the figure below.



CILO = cilofexor; OLE = open-label extension; PTM = placebo to match; R/O = rollover.

For both the Blinded Study Phase and OLE Phase, at the discretion of the principal investigator (PI), study drug dosing may be temporarily interrupted, for example, due to an adverse event (AE). During the period of study drug dosing interruption, participants should continue with study visits per the Study Procedures Table (Appendix 1). Study drug dosing may be reinitiated at the discretion of the PI. Dose interruption/reduction for the management of pruritus (see Section 7.5.3 of the protocol) or for other AEs (see Section 7.5.4 of the protocol) is permitted.

Participants and all personnel directly involved in the conduct of the study will be blinded to treatment assignment throughout the Blinded Study Phase. After all participants have completed Blinded Study Phase Week 96 or ET, the treatment assignments will be unblinded to the sponsor only. Investigators and participants will remain blinded to the Blinded Study Phase treatment assignment until approximately 6 weeks after all participants have completed the OLE Phase follow-up visit. Refer to Section 5.1.2 of the protocol for more details.

1.3. Sample Size and Power

A sample size of 267 participants in the CILO group and 133 participants in the placebo group has 81% power to detect an absolute difference of 15% in the percentage of participants who meet the primary endpoint at Week 96. Power was calculated using Pearson's Chi-square test at a 2-sided significance level of 0.05. This calculation assumes that 25% of participants will discontinue the study prematurely (considered as treatment failures), and that among participants with nonmissing response data at Blinded Study Phase Week 96, 20% in the CILO group and 40% in the placebo group will meet the primary endpoint.

The primary endpoint rate of 40% for participants in the placebo group with nonmissing response data at Blinded Study Phase Week 96 was estimated based on Week 96 data from noncirrhotic participants in the simtuzumab Phase 2 study (GS-US-321-0102).

2. TYPE OF PLANNED ANALYSIS

2.1. Interim Analyses

2.1.1. DMC Analysis

An external multidisciplinary Data Monitoring Committee (DMC) will review the progress of the study and perform interim reviews of the safety data in order to protect participant welfare and preserve study integrity. To ensure the best interests of the participants, the DMC will make recommendations to the sponsor if the nature, frequency, and severity of adverse effects associated with the study treatment warrant the early termination of the study, the continuation of the study, or the continuation of the study with modifications.

The initial DMC data review meeting will be conducted after approximately 50 participants complete Week 4 in the study. Additional meetings will be scheduled every 6 months following the initial meeting. The frequency of the meetings can be adjusted if deemed necessary by the DMC. Documentation of such change may be done in the meeting minutes or a note to file without an amendment to the charter.

The DMC's role and responsibilities and the scope of analysis to be provided to the DMC are provided in a mutually agreed upon charter, which defines the DMC membership, meeting logistics, and meeting frequency.

2.1.2. Planned Interim Futility Analysis

There will be one planned interim futility analysis based on the primary endpoint after the first 160 randomized and dosed participants have completed Week 96 or ET assessments in the Blinded Study Phase. A nonbinding futility rule based on predictive power (PP) approach will be used for futility assessment. The DMC may recommend early termination of the study due to futility if the criterion of PP $\leq$ 10% is met.

The PP is calculated by assuming a noninformative prior Beta(1, 1) for the response rate in each treatment group and each stratum, updating it to a posterior distribution using observed data at the interim futility analysis, and then computing the averaged conditional power over the posterior distributions, where the conditional power is the probability of successfully rejecting the null hypothesis at the primary analysis as described in Section 6.1.4, given the data observed at the interim futility analysis and an assumed response rate in each treatment group and each stratum. See Appendix 4 for detailed calculations of PP.

In addition to the above described futility analysis based on the primary endpoint, selected analysis of efficacy endpoints described in Section 6 will also be included in the interim futility analysis.

2.2. Primary Analysis

The unblinded primary analyses of the efficacy endpoints (ie, primary and secondary) at Blinded Study Phase Week 96 will be conducted as described in Section 6. The study blind will be broken and the analysis will be conducted after all participants have completed or prematurely discontinued from the Blinded Study Phase of the study and completed the follow-up visit, if applicable, for the Blinded Study Phase, outstanding data queries have been resolved or adjudicated as unresolvable, and the data have been cleaned and finalized for the analysis.

2.3. Final Analysis

The final analysis will be performed after all participants have completed the OLE Phase follow-up visit or prematurely discontinued from the study, outstanding data queries have been resolved or adjudicated as unresolvable, and the data have been cleaned and finalized.

3. GENERAL CONSIDERATIONS FOR DATA ANALYSES

Analysis results will be presented using descriptive statistics. For categorical variables, the number and percentage of participants in each category will be presented; for continuous variables, the number of participants (n), mean, standard deviation (SD) or standard error (SE), median, first quartile (Q1), third quartile (Q3), minimum, and maximum will be presented.

All statistical tests will be 1-sided and performed at the 2.5% significance level unless otherwise specified.

By-participant listings will be presented for all participants in the All Randomized Analysis Set and sorted by participant identification (ID) number, visit date, and time (if applicable). Data collected on log forms, such as AEs, will be presented in chronological order for each participant. The treatment group to which participants were initially assigned will be used in the listings. Age, sex at birth, race, and ethnicity will be included in the listings, as space permits.

3.1. Analysis Sets

Analysis sets define the participants to be included in an analysis. Analysis sets and their definitions are provided in this section. Participants included in each analysis set will be determined before the study blind is broken for analysis. The analysis set will be identified and included as a subtitle of each table, figure, and listing.

A listing of reasons for exclusion from analysis sets will be provided by participants.

3.1.1. All Randomized Analysis Set

All Randomized Analysis Set includes all participants who were randomized in the study.

3.1.2. Full Analysis Set

The Full Analysis Set (FAS) includes all randomized participants who took at least 1 dose of study drug. This is the primary analysis set for efficacy analyses.

3.1.3. Safety Analysis Set

The Safety Analysis Set includes all participants who took at least 1 dose of study drug. This is the primary analysis set for safety analyses.

3.1.4. OLE Analysis Set

The OLE Analysis Set includes all participants who took at least 1 dose of study drug in the OLE Phase. This is the primary analysis set for safety and efficacy analyses for the OLE Phase.

3.1.5. Pharmacokinetic Analysis Set

The PK Analysis Set will include all randomized participants who took at least 1 dose of study drug and have at least 1 nonmissing concentration of CILO (and its metabolites, as applicable) reported by the bioanalytical laboratory. This is the primary analysis set for all PK analyses.

CCI [redacted]

CCI [redacted]

3.2. Participant Grouping

For analyses based on the FAS, participants will be grouped according to the treatment to which they were randomized.

For analyses based on the Safety Analysis Set and PK Analysis Set, participants will be grouped according to the actual treatment received. The actual treatment received will differ from the randomized treatment only when their actual treatment differs from randomized treatment for the entire treatment duration in the Blinded Study Phase. In this case, the actual treatment received is defined as the treatment received for the entire treatment duration in the Blinded Study Phase. The treatment mentioned in this section refers to CILO or placebo, regardless of the dose level.

For analyses in the Blinded Study Phase, participants will be grouped into the following Analysis Groups, and only the Blinded Study Phase data will be summarized:

1) Blinded Phase CILO: this group includes all participants who were on CILO 100 mg in the Blinded Study Phase.

2) Blinded Phase Placebo: this group includes all participants who were on placebo in the Blinded Study Phase.

For analyses in the OLE Phase, participants will be grouped into the following Analysis Groups, and only the OLE Phase data will be summarized:

3) OLE Phase CILO - Previously on CILO: this group includes all participants who received OLE study drug and previously received CILO 100 mg in the Blinded Study Phase.

4) OLE Phase CILO - Previously on Placebo: this group includes all participants who received OLE study drug and previously received placebo in the Blinded Study Phase.

3.3. Strata and Covariates

Participants will be randomly assigned to treatment groups via an interactive response technology (IRT) in a 2:1 ratio using a stratified randomization schedule. Stratification will be based on the following variables:

- UDCA use: yes or no
- Fibrosis stage according to Ludwig classification by central reader: F3 or F0-F2

If there are discrepancies in stratification factor values between the IRT and the clinical database, the values recorded in the clinical database will be used for analyses.

Efficacy endpoints will be evaluated using stratification factors as specified in Section 6.

3.4. Examination of Participant Subgroups

Subgrouping of participants based on randomization stratification factors will be explored for subgroup analyses. If there is an imbalance between treatment groups in presumed prognostic baseline characteristics that are not stratification factors, subgroupings based on these imbalanced baseline characteristics will also be explored for analysis of the primary endpoint. The presumed prognostic baseline characteristics (defined based on nonmissing values) include the following:

- UDCA use (yes and no)
- Fibrosis stage (F3 and F0-F2 according to the Ludwig classification)
- ALP ($\leq 1.5 \times$ upper limit of normal [ULN] and $> 1.5 \times$ ULN)
- ELF score (< 9.8 and ≥ 9.8)
- Sex (male and female)
- Magnetic resonance cholangiopancreatography (MRCP) (intrahepatic duct involvement only, both intra- and extrahepatic duct involvement, and intrahepatic duct involvement only or both intra- and extrahepatic duct involvement)
- Inflammatory bowel disease (IBD) history (yes and no)

For the comparison of treatment difference in each subgroup, the point estimate and 95% confidence interval (CI) based on the primary analysis approach as described in Section 6.1.4 will be presented.

3.5. Multiple Comparisons

The study has one primary efficacy endpoint: the proportion of participants with progression of liver fibrosis, as defined by a ≥ 1-stage increase in fibrosis according to the Ludwig classification score at Blinded Study Phase Week 96. The primary efficacy endpoint will be tested at the significance level of 0.025 (1-sided).

The family-wise type I error rate for the primary and the secondary efficacy endpoints comparisons will be controlled at 2.5% (1-sided). If the primary efficacy endpoint null hypothesis is rejected, the null hypotheses for the secondary endpoints will be tested in the order presented below. Each test will be performed at the same 1-sided 2.5% significance level. The formal hypothesis test for the subsequent endpoint(s) will stop after the first failure to reject the null hypothesis of no treatment benefit for an endpoint.

1. Change from baseline in serum ALP at Blinded Study Phase Week 96
2. Change from baseline in serum ALT at Blinded Study Phase Week 96
3. Change from baseline in serum bile acids at Blinded Study Phase Week 96
4. The proportion of participants with ≥ 25% relative reduction in serum ALP concentration from baseline (biochemical response) and no worsening of fibrosis according to the Ludwig classification (histologic response) at Blinded Study Phase Week 96
5. The proportion of participants with fibrosis improvement (according to the Ludwig classification) at Blinded Study Phase Week 96
6. Change from baseline in PSC Symptoms – Module 1 on the disease-specific PSC-PRO at Blinded Study Phase Week 96
7. Change from baseline in ELF test score at Blinded Study Phase Week 96
8. Change from baseline in liver stiffness by FibroScan at Blinded Study Phase Week 96

CCI [redacted].

3.6. Missing Data and Outliers

3.6.1. Missing Data

For missing last dosing date of study drug, imputation rules are described in Section 4.2.1. The handling of missing or incomplete dates for AE onset is described in Section 7.1.5, and for prior and concomitant medications in Section 7.4.

The following imputation approaches will be applied to efficacy endpoints as specified:

1) Non-Responder Imputation (NRI)

   For purpose of analysis at a defined time point, when NRI is used, all participants with missing values will be analyzed as treatment failures. The NRI method will be applied to participants with or without baseline values.

2) Multiple Imputation (MI)

   The MI procedure replaces each missing data point with a set of plausible values that represent the uncertainty about the right value to impute under the assumption of missing at random (MAR). The imputation model used to impute missing data points will be based on a logistic regression or linear regression (depending on the type of the endpoint) with the treatment groups and the 2 stratification factors (and other predictor variables as appropriate) as covariates. A total of 50 imputed datasets will be generated by the imputation model. That is, each missing data point will be imputed 50 times. Each imputed data set will then be analyzed by the method for the primary and secondary analyses as specified in Section 6.1.4 and 6.2.3. The results from these 50 imputed data sets will then be combined using Rubin's rule {Rubin 1987}.

3) Tipping Point Analysis (TPA)

   If any significant result in favor of CILO 100 mg is detected in the primary analysis of the binary primary efficacy endpoint (described in Section 6.1.4), a 2-dimensional delta-adjusting pattern-mixture approach for tipping point analysis {Ratitch 2013} will be conducted to assess the robustness of the primary analysis result under missing not at random (MNAR) assumption for the CILO and placebo groups. Specifically, it is assumed that a systematic difference exists between the conditional distributions of the missing and observed data in each group. To reflect such a systematic difference, a shift parameter δ_j will be applied to the imputation model (ie, the logistic regression with the treatment groups and the 2 stratifications factors as covariates) when the missing data points in each group are imputed, where $j = A,B$, representing the CILO and placebo groups, respectively. The values of δ_A will be unfavorable for the CILO group and the values of δ_B will be favorable for the placebo group, ie, increasing the probability of being a responder of fibrosis progression for CILO participants and reducing the probability of being a responder for placebo participants. For each paired values of (δ_A,δ_B), multiple imputed data sets will be generated, each data set will be analyzed by the same method for the primary analysis, and analysis results will then be combined by Rubin's rule. Thus, by varying the values of (δ_A,δ_B), the impact from missing data on the analysis results will be examined so as to identify the tipping point (ie, the values at which conclusions from statistical inference change from being significant to being insignificant for evaluation of CILO over placebo).

4) Last Observation Carried Forward (LOCF)

   For purpose of analysis at a defined time point, if the value is missing, the last observed value prior to the time point will be used to impute the missing value. 'Last observation' includes postbaseline records, and only the records eligible for efficacy analysis will be

carried forward. Refer to Section 3.8.2 for the handling of the data to be included in efficacy analysis. LOCF will be applied to those participants who have a baseline value and at least one postbaseline value at an analysis visit. No imputation will be performed if the value at baseline is missing. Participants with missing baseline values or missing Week 96 values after applying the LOCF method will be excluded from the corresponding analysis.

5) Observed Case (OC)

Only participants with observed Baseline and Week 96 values will be included in the corresponding analysis.

The NRI approach will be used for the primary analysis of the primary efficacy endpoint and secondary binary endpoints. The MI approach will be used for the main analysis of the secondary continuous endpoints.

The MI and TPA approaches will be used for sensitivity analyses of the primary efficacy endpoint. LOCF analysis and OC analysis will be carried out as sensitivity analyses for all primary and secondary efficacy endpoints using the same analysis method as defined in Sections 6.1.4 and 6.2.3.

For the primary and secondary efficacy endpoints, all the imputation approaches will be applied under the estimand framework as defined in Sections 6.1.2 and 6.2.2.

CCI [redacted]

3.6.2. Outliers

Outliers will be identified during the data management and data analysis process. All data, including outliers, will be included in analyses. If necessary, selected endpoints may be evaluated using a rank-based analysis method to minimize the influence of outliers.

3.7. Data Handling Conventions and Transformations

The following conventions will be used for the imputation of date of birth when it is partially missing or not collected:

- If only month and year of birth is collected, then "15" will be imputed as the day of birth
- If only year of birth is collected, then "01 July" will be imputed as the day and month of birth
- If year of birth is missing, then date of birth will not be imputed.

In general, age collected at Day 1 (in years) will be used for analyses and presented in listings. If age at Day 1 is not available for a participant, then age derived based on date of birth and the Day 1 visit date will be used instead. If an enrolled participant was not dosed with any study drug, the randomization date will be used instead of the Day 1 visit date. For screen failures, the

date of the first informed consent from the last screen (if the participant was rescreened) was signed will be used for the age derivation. Age required for longitudinal and temporal calculations and analyses (e.g., estimates of creatinine clearance, age at date of AE) will be based on age derived from date of birth and the date of the measurement or event, unless otherwise specified.

Non-PK data that are continuous in nature but are less than the lower limit of quantitation (LOQ) or above the upper LOQ will be imputed as follows:

- A value that is 1 unit less than the LOQ will be used to calculate descriptive statistics if the datum is reported in the form of "< x" (where x is considered the LOQ). For example, if the values are reported as < 50 and < 5.0, values of 49 and 4.9, respectively, will be used to calculate summary statistics. An exception to this rule is any value reported as < 1 or < 0.1, etc. For values reported as < 1 or < 0.1, a value of 0.9 or 0.09, respectively, will be used to calculate summary statistics.

- A value that is 1 unit above the LOQ will be used to calculate descriptive statistics if the datum is reported in the form of "> x" (where x is considered the LOQ). Values with decimal points will follow the same logic as above.

- The lower or upper LOQ will be used to calculate descriptive statistics if the datum is reported in the form of "≤ x" or "≥ x" (where x is considered the lower or upper LOQ, respectively).

When any ELF component (hyaluronic acid, procollagen 3 N-terminal propeptide [P3NP], and tissue inhibitor of metalloproteinase 1 [TIMP1]) is less than the lower LOQ or above the upper LOQ, the ELF component(s) will be imputed based on the rules described above, and the ELF test score will be calculated based on the imputed value of the component(s) as defined in Appendix 3.

If methods based on the assumption that the data are normally distributed are not adequate, analyses may be performed on transformed data or nonparametric analysis methods may be used, as appropriate.

Sparse PK concentration values that are below the limit of quantitation (BLQ) will be presented as "BLQ" in the data listing.

Natural logarithm transformation will be used for analyzing concentrations and PK parameters in intensive PK samples. Concentration values that are BLQ will be presented as "BLQ" in the concentration data listing. Values that are BLQ will be treated as 0 at predose time points, and one-half the value of the LOQ at postdose time points for summary purposes.

The following conventions will be used for the presentation of summary and order statistics for intensive PK concentrations:

- If at least 1 participant has a concentration value of BLQ for the time point, the minimum value will be displayed as "BLQ."
- If more than 25% of the participants have a concentration data value of BLQ for a given time point, the minimum and Q1 values will be displayed as "BLQ."
- If more than 50% of the participants have a concentration data value of BLQ for a given time point, the minimum, Q1, and median values will be displayed as "BLQ."
- If more than 75% of the participants have a concentration data value of BLQ for a given time point, the minimum, Q1, median, and Q3 values will be displayed as "BLQ."
- If all participants have concentration data values of BLQ for a given time point, all order statistics (minimum, Q1, median, Q3, and maximum) will be displayed as "BLQ."

PK parameters that are BLQ will be imputed as one-half LOQ before log transformation or statistical model fitting.

3.8. Analysis Visit Windows

3.8.1. Definition of Study Day

Study day/OLE study day will be calculated from the first dosing date of study drug in the Blinded Study Phase/OLE Phase and derived as follows:

- For postdose study days: Assessment Date – First Dosing Date of the study phase + 1
- For days prior to the first dose: Assessment Date – First Dosing Date of the study phase

Therefore, study day 1/OLE study day 1 is the day of first dose of study drug administration in the corresponding study phase.

3.8.2. Analysis Visit Windows

Participant visits might not occur on protocol-specified days. Therefore, for the purpose of analysis, observations will be assigned to analysis windows for each of the 2 reporting phases: Blinded Study Phase and OLE Phase.

On-treatment visit windows will be calculated from Day 1 for selected efficacy measures, vital signs, elastography, and safety laboratory data of the respective reporting phase.

For Blinded Study Phase, except for histologic biopsy data, MRCP, and liver stiffness by transient elastography data, selected safety and efficacy data (unless otherwise specified) collected up to and including the last dosing date + 30 days and before the first dose of OLE Phase for participants who have permanently discontinued study drug in Blinded Study Phase, or the database snapshot date (or data cut date if applicable) for participants who were still on treatment in Blinded Study Phase at the time of interim analysis, will be mapped according to the following analysis windows unless the nominal visit name is Follow-Up.

The analysis windows for selected measures to be applied to the Blinded Study Phase are provided in Table 3-1 to Table 3-10. The algorithm for assigning baseline analysis window does not apply to the liver tests (ALP, total bilirubin, ALT, aspartate aminotransferase [AST], and GGT). For these 5 parameters, the baseline values will be determined by averaging the values obtained between Screening and Baseline/Day 1.

Table 3-1. Analysis Visit Windows in Blinded Study Phase for Chemistry, Hematology, Coagulation, Child-Pugh (CP), model for end-stage liver disease (MELD) scores, Vital Signs, Body Weight, Partial Mayo Score, Pruritus VAS, 5D-Itch, Mayo Risk Score and Amsterdam-Oxford Score

Nominal Visit	Nominal Study Day	Visit Window Study Day	
		Lower Limit	Upper Limit
Baseline	1	(none)	1
Week 4	29	2	42
Week 8	57	43	70
Week 12	85	71	126
Week 24	169	127	210
Week 36	253	211	294
Week 48	337	295	378
Week 60	421	379	462
Week 72	505	463	546
Week 84	589	547	630
Week 96	673	631	≥ 673

Table 3-2. Analysis Visit Windows in Blinded Study Phase for Total Bile Acids

Nominal Visit	Nominal Study Day	Visit Window Study Day	
		Lower Limit	Upper Limit
Baseline	1	(none)	1
Week 4	29	2	56
Week 12	85	57	126
Week 24	169	127	210
Week 36	253	211	294
Week 48	337	295	378
Week 60	421	379	462
Week 72	505	463	588
Week 96	673	589	≥ 673

Table 3-3. Analysis Visit Windows in Blinded Study Phase for Lipid Profile

Nominal Visit	Nominal Study Day	Visit Window Study Day	
		Lower Limit	Upper Limit
Baseline	1	(none)	1
Week 12	85	2	126
Week 24	169	127	210
Week 36	253	211	294
Week 48	337	295	378
Week 60	421	379	462
Week 72	505	463	546
Week 84	589	547	630
Week 96	673	631	≥ 673

Table 3-4. Analysis Visit Windows in Blinded Study Phase for C4 and FGF19

Nominal Visit	Nominal Study Day	Visit Window Study Day	
		Lower Limit	Upper Limit
Baseline	1	(none)	1
Week 12	85	2	126
Week 24	169	127	252
Week 48	337	253	420
Week 72	505	421	588
Week 96	673	589	≥ 673

Table 3-5. Analysis Visit Windows in Blinded Study Phase for C-Peptide, Insulin and HbA1c

Nominal Visit	Nominal Study Day	Visit Window Study Day	
		Lower Limit	Upper Limit
Baseline	1	(none)	1
Week 12	85	2	126
Week 24	169	127	252
Week 48	337	253	504
Week 96	673	505	≥ 673

Table 3-6. Analysis Visit Windows in Blinded Study Phase for ELF Test Score and Components, FibroScan, Health Resource Utilization, and QoL

Nominal Visit	Nominal Study Day	Visit Window Study Day	
		Lower Limit	Upper Limit
Baseline	1	(none)	1
Week 24	169	2	252
Week 48	337	253	420
Week 72	505	421	588
Week 96	673	589	≥ 673

Table 3-7. Analysis Visit Windows in Blinded Study Phase for C-Reactive Protein (CRP)

Nominal Visit	Nominal Study Day	Visit Window Study Day	
		Lower Limit	Upper Limit
Baseline	1	(none)	1
Week 24	169	2	252
Week 48	337	253	504
Week 96	673	505	≥ 673

Table 3-8. Analysis Visit Windows in Blinded Study Phase for FibroSURE/FibroTest and Selected Components (α2-Macroglobin, Haptoglobin, Apolipoprotein A1)

Nominal Visit	Nominal Study Day	Visit Window Study Day	
		Lower Limit	Upper Limit
Baseline	1	(none)	1
Week 48	337	2	504
Week 96	673	505	≥ 673

Table 3-9. Analysis Visit Windows in Blinded Study Phase for MRCP

Nominal Visit	Nominal Study Day	Visit Window Study Day	
		Lower Limit	Upper Limit
Baseline	1	(none)	57
Week 48	337	58	504
Week 96	673	505	≥ 673

Table 3-10. Analysis Visit Windows in Blinded Study Phase for Liver Biopsy

Nominal Visit	Nominal Study Day	Visit Window Study Day	
		Lower Limit	Upper Limit
Baseline	1	(none)	1
Week 96	673	337	≥ 673

For OLE Phase, except for liver stiffness by transient elastography data, selected safety and efficacy data (unless otherwise specified) collected up to and including the last dosing date + 30 days for participants who have permanently discontinued study drug in OLE Phase, or the database snapshot date (or data cut date if applicable) for participants who were still on treatment in OLE Phase at the time of interim analysis or primary analysis, will be mapped according to the following analysis windows unless the nominal visit name is OLE Follow-Up.

The analysis windows for selected measures to be applied to the OLE Phase are provided in Table 3-11 to Table 3-13. The algorithm for assigning OLE baseline analysis window does not apply to the liver tests (ALP, total bilirubin, ALT, AST, and GGT). Among participants whose duration of time since the last dosing date in the Blinded Study Phase is 28 days or longer, the OLE baseline value of liver tests will be determined by averaging all available values that are between 28 days prior to and inclusive of OLE Day 1. All laboratory values will be included in the averaging, regardless of scheduled or unscheduled laboratory collection. For participants initiating OLE Day 1 less than 28 days from the last dosing date of the Blinded Study Phase, the OLE baseline value will be the last available value on or prior to OLE Day 1.

Table 3-11. Analysis Visit Windows in OLE Phase for Chemistry, Hematology, Coagulation, CP, MELD scores, Vital Signs, Body Weight, Partial Mayo Score, Pruritus VAS, 5D-Itch, Mayo Risk Score and Amsterdam-Oxford Score

Nominal Visit	Nominal Study Day	Visit Window Study Day	
		Lower Limit	Upper Limit
OLE Baseline	1	(none)	1
OLE Week 4	29	2	98
OLE Week 24	169	99	252
OLE Week 48	337	253	420
OLE Week 72	505	421	588
OLE Week 96	673	589	≥ 673

Table 3-12. Analysis Visit Windows in OLE Phase for Lipid Profile, FibroScan, Health Resource Utilization, and QoL

Nominal Visit	Nominal Study Day	Visit Window Study Day	
		Lower Limit	Upper Limit
OLE Baseline	1	(none)	1
OLE Week 24	169	2	252
OLE Week 48	337	253	420
OLE Week 72	505	421	588
OLE Week 96	673	589	≥ 673

Table 3-13. Analysis Visit Windows in OLE Phase for Total Bile Acids, CRP, ELF Test Score and Components, FibroSURE/FibroTest and Selected Components (α2-Macroglobin, Haptoglobin, Apolipoprotein A1), C4 and FGF19

Nominal Visit	Nominal Study Day	Visit Window Study Day	
		Lower Limit	Upper Limit
OLE Baseline	1	(none)	1
OLE Week 48	337	2	504
OLE Week 96	673	505	≥ 673

Data relating to unscheduled visits and early termination visits may be assigned to a particular visit or time point based on the analysis visit windows. The following conventions will be followed:

- An unscheduled visit prior to the first dosing of study drug may be included in the calculation of the baseline value, if applicable.
- Unscheduled visits after the first dose of study drug will be included in determining the maximum postbaseline toxicity grade.
- Data collected on a follow-up visit in the Blinded Study Phase will be summarized as a separate visit, and labeled "Follow-up Visit." The analysis window for the Follow-up Visit includes the corresponding nominal visit and all unscheduled visits occurred after but prior to the OLE Day 1 visit.
- Data collected on an early termination follow-up visit in the Blinded Study Phase will be summarized as a separate visit, and labeled "Early Termination Follow-up Visit." The analysis window for the Early Termination Follow-up Visit includes the corresponding nominal visit and all unscheduled visits occurred after but prior to the OLE Day 1 visit.
- Data collected on a follow-up visit in the OLE Phase will be summarized as a separate visit, and labeled "OLE Follow-up Visit." The analysis window for the OLE Follow-up Visit includes the corresponding nominal visit and all unscheduled visits occurred after.
- Data collected on an early termination follow-up visit in the OLE Phase will be summarized as a separate visit, and labeled "OLE Early Termination Follow-up Visit." The analysis window for the OLE Early Termination Follow-up Visit includes the corresponding nominal visit and all unscheduled visits occurred after.
- Except for histologic biopsy data, MRCP, and liver stiffness by transient elastography data, selected safety and efficacy data obtained after the last dose date plus 30 days for participants who have permanently discontinued study drug in each study phase will be excluded from the summaries, but will be included in the listings. For participants entering the OLE Phase, the Blinded Study Phase ends on the day before OLE Day 1.

3.8.3. Selection of Data in the Event of Multiple Records in an Analysis Visit Window

Depending on the statistical analysis method, single value may be required for each analysis window. For example, change from baseline by visit usually requires a single value, whereas a time-to-event analysis would not require 1 value per analysis window.

If multiple valid, nonmissing measurements exist in an analysis window, records will be chosen based on the following rules if a single value is needed:

- In general, the baseline value for Blinded Study Phase and OLE Phase will be the last nonmissing value on or prior to the first dosing date of study drug of the respective study phase, unless specified differently. If there are multiple records with the same time or no time recorded on the same day, the baseline value will be the average of the measurements for continuous data, or the measurement with the lowest severity (e.g., normal will be selected over abnormal for safety electrocardiogram [ECG] findings) for categorical data.
- For each liver test (ALP, total bilirubin, ALT, AST, and GGT), the baseline value in the Blinded Study Phase will be determined by averaging the values obtained between Screening from the last screen (if the participant was rescreened) and Baseline/Day 1. The corresponding baseline reference range will be defined as the one associated with the latest visit that was included for computing the baseline value, for the purpose of determination of the abnormality and/or toxicity grades.
- For each liver test (ALP, total bilirubin, ALT, AST, and GGT), among participants whose duration of time since the last dosing date in the Blinded Study Phase is 28 days or longer, the OLE baseline value of liver tests will be determined by averaging all available values that are between 28 days prior to and inclusive of OLE Day 1. All laboratory values will be included in the averaging, regardless of scheduled or unscheduled laboratory collection. For participants initiating OLE Day 1 less than 28 days from the last dosing date of the Blinded Study Phase, the OLE baseline value will be the last available value on or prior to OLE Day 1. The corresponding baseline reference range will be defined as the one associated with the latest visit that was included for computing the baseline value, for the purpose of determination of the abnormality and/or toxicity grades.
- For postbaseline values:
  - The record closest to the nominal day for that visit will be selected.
  - If there are 2 or more records that are equidistant from the nominal day, the later record will be selected.
  - If there is more than 1 record on the selected day, the average will be taken for continuous data and the worst severity will be taken for categorical data, unless otherwise specified.

Liver stiffness by transient elastography data in each analysis visit window will be chosen based on the following rules for each reporting phase:

- For baseline, the last nonmissing value on or prior to the first dosing date of study drug will be selected. If multiple measurements by different probe types (XL or M) are available on the same day, the measurement by M probe will be selected for analysis.
- For postbaseline visits, measurements by the same probe type selected for the participant at baseline will be selected in each analysis visit window. If no measurement by the same probe type as baseline is available, the measurements by the other probe type analysis value for the corresponding postbaseline visit will be considered. If multiple postbaseline records by the same probe type are available, the rules to choose postbaseline continuous measurements as described above will apply.

4. PARTICIPANT DISPOSITION

4.1. Participant Enrollment and Disposition

Key study dates (ie, first participant screened, first participant randomized, last participant randomized, last participant last visit for the primary endpoint, and last participant last visit for the clinical study report) will be provided.

A summary of participant enrollment will be provided by treatment group and overall for each country and investigator within a country. The summary will present the number and percentage of participants enrolled. For each column, the denominator for the percentage calculation will be the total number of participants analyzed for that column.

A similar enrollment table will be provided by randomization stratum. The denominator for the percentage of participants in the stratum will be the total number of enrolled participants. If there are discrepancies in the value used for stratification assignment between the IRT and the clinical database, the value collected in the clinical database will be used for the summary. A listing of participants with discrepancies in the value used for stratification assignment between the IRT and the clinical database at the time of data finalization will be provided.

The randomization schedule used for the study will be provided as an appendix to the CSR.

A summary of participant disposition will be provided by treatment group and study phase and overall. This summary will present the number of participants screened, the number of participants who met all eligibility criteria but were not randomized with reasons participants not randomized, the number of participants randomized, and the number of participants in each of the categories listed below (as applicable):

- Randomized but never treated
- Safety Analysis Set
- Full Analysis Set
- Pharmacokinetic Analysis Set
- [redacted]
- Study drug completion status in Blinded Study Phase and reasons for premature discontinuation
- OLE Analysis Set
- Study drug completion status in OLE Phase and reasons for premature discontinuation
- Study completion status and reasons for premature discontinuation

For the status of study drug and study completion and reasons for premature discontinuation, the number and percentage of participants in each category will be provided. The denominator for the percentage calculation will be the total number of participants in the Safety Analysis Set for Blinded Study Phase and OLE analysis set for OLE Phase corresponding to that column. In addition, a flowchart will be provided to depict the disposition.

The following by-participant listings will be provided by participant identification (ID) number in ascending order to support the above summary tables:

- Reasons for premature study drug or study discontinuation
- Reasons for screen failure (will be provided by screening ID number in ascending order)
- Lot number and kit ID

4.2. Extent of Study Drug Exposure and Adherence

Extent of exposure to study drug will be examined by assessing the total duration of exposure to study drug and the level of adherence relative to the study drug regimen specified in the protocol.

4.2.1. Duration of Exposure to Study Drug

Total duration of exposure to study drug in Blinded Study Phase and OLE Phase, respectively, will be defined as last dosing date minus first dosing date plus 1, regardless of any temporary interruptions in study drug administration, and will be expressed in weeks using up to 1 decimal place (eg, 4.5 weeks). If the last study drug dosing date is missing, the latest date among the study drug end date, clinical visit date, laboratory sample collection date, and vital signs assessment date that occurred during the on-treatment period for participants who have permanently discontinued study drug or the database snapshot date (or data cut date if applicable) for participants who were still on treatment at the time of the interim analysis or primary analysis will be used. If month and year of the last dose are known, and the last study drug dosing date imputed above is different from the month collected, the last date of that month will be used. If only year of the last dose is known, and the last study drug dosing date imputed above is after the year collected, the last date of that year will be used; if the last study drug dosing date imputed above is before the year collected, the first date of that year will be used.

The total duration of exposure to study drug will be summarized using descriptive statistics and using the number (ie, cumulative counts) and percentage of participants exposed through the following time periods:

For Blinded Study Phase: 1 day, 4 weeks, 8 weeks, 12 weeks, and every 12 weeks thereafter in the Blinded Study Phase. Summaries will be provided by Analysis Groups 1 and 2 using the Safety Analysis Set.

For OLE Phase: 1 day, 4 weeks, 24 weeks, and every 24 weeks thereafter in the OLE Phase. Summaries will be provided by Analysis Groups 3 and 4 using the OLE Analysis Set.

The analysis sets and Analysis Groups are defined in Sections 3.1 and 3.2.

No formal statistical testing is planned.

4.2.2. Adherence to Study Drug

The total number of doses (in the unit of tablets) administered will be summarized using descriptive statistics.

The presumed total number of doses administered to a participant will be determined by the data collected on the drug accountability eCRF using the following formula:

Total Number of Doses Administered =

$$\left(\sum \text{No. of Doses Dispensed}\right) - \left(\sum \text{No. of Doses Returned}\right)$$

4.2.2.1. On-Treatment Adherence

On-treatment adherence rate will be calculated for the Blinded Study Phase and OLE Phase and will be provided by Analysis Groups 1 and 2 using the Safety Analysis Set and by Analysis Groups 3 and 4 (Section 3.2) using the OLE Analysis Set. The level of on-treatment adherence to the study drug regimen will be determined by the total amount of study drug administered relative to the total amount of study drug expected to be administered during a participant's actual on-treatment period based on the study drug regimen.

The level of on-treatment adherence will be expressed as a percentage using the following formula:

$$\text{On-Treatment Adherence (\%)} = \left(\frac{\text{Total Amount of Study Drug Administered}}{\text{Study Drug Expected to be Administered on Treatment}}\right) \text{x } 100$$

Descriptive statistics for the level of on-treatment adherence with the number and percentage of participants belonging to adherence categories (eg, $< 75\%$, ≥ 75 to $< 90\%$, $\geq 90\%$) will be provided by Analysis Group for the Blinded Study Phase and OLE Phase, respectively.

No formal statistical testing is planned.

A by-participant listing of study drug administration and drug accountability will be provided separately by participant ID number (in ascending order) and visit (in chronological order).

4.3. Protocol Deviations

Participants who did not meet the eligibility criteria for study entry, but enrolled in the study will be summarized regardless of whether they were exempted by the sponsor or not. The summary will present the number and percentage of participants who did not meet at least 1 eligibility criterion and the number of participants who did not meet specific criteria by treatment group based on All Randomized Analysis Set. A by-participant listing will be provided for those participants who did not meet at least 1 eligibility (inclusion or exclusion) criterion. The listing will present the eligibility criterion (or criteria if more than 1 deviation) that participants did not meet and related comments, if collected.

Protocol deviations occurring after participants entered the study are documented during routine monitoring. The number and percentage of participants with important protocol deviations (IPDs) by deviation reason (eg, eligibility criteria, informed consent) will be summarized by Analysis Group for the Blinded Study Phase using the All Randomized Analysis Set and for the OLE Phase using the OLE Analysis Set. A by-participant listing will be provided for those participants with important protocol deviations.

4.4. Assessment of COVID-19 Impact

This study was ongoing during the novel coronavirus (COVID-19) pandemic which has an impact on the study conduct. Some participants were unable to attend onsite visits due to shelter in place guidelines, site closures, or other reasons. This section describes how special situations due to COVID-19 will be handled in the analysis.

4.4.1. Study Drug or Study Discontinuation Due to COVID-19

A by-participant listing of reasons for premature study drug or study discontinuation due to COVID-19 will be provided if applicable.

4.4.2. Protocol Deviations Due to COVID-19

A summary of IPDs due to COVID-19 will be provided for the Blinded Study Phase and OLE Phase, similar to the summary described in the protocol deviations section (Section 4.3).

The number and percentage of participants with nonimportant protocol deviations related to COVID-19 by number of deviations (eg, at least 1, with 1, 2, 3, or more deviations) will be summarized by Analysis Group for the Blinded Study Phase using the All Randomized Analysis Set and for the OLE Phase using the OLE Analysis Set.

A by-participant listing will be provided for participants with IPDs related to COVID-19 if applicable. A separate listing will be provided for participants with nonimportant protocol deviations related to COVID-19 if applicable.

4.4.3. Missed and Virtual Visits due to COVID-19

A summary of participants affected by COVID-19 pandemic will be provided for each scheduled study visit by Analysis Group and overall for the Blinded Study Phase using the Safety Analysis Set and for the OLE Phase using the OLE Analysis Set. For each visit, the summary will present the number and percentage of participants who missed the visit due to COVID-19 or had a virtual visit due to COVID-19. For each column, the denominator for the percentage calculation will be the number of participants at risk for that column and the given visit.

A by-participant listing of participants with missed or virtual visits due to COVID-19 will be provided by participant ID number in ascending order.

Information regarding missed or virtual visits due to COVID-19 will be collected as free text in the CRF comment fields. The determination of missed or virtual visits due to COVID-19 will be done using Natural Language Processing (NLP) to search the CRF comment fields. A detailed explanation of the algorithm is given in Appendix 5.

4.4.4. Adverse Events Due to COVID-19

Adverse events (AEs) of COVID-19 will be included in analyses of AEs for the Blinded Study Phase using the Safety Analysis Set and for the OLE Phase using the OLE Analysis Set if applicable, which will be determined through COVID-19 SMQ narrow search. A by-participant listing of AEs of COVID-19 will be provided if applicable.

4.4.5. Overall Assessment of COVID-19 Pandemic Impact

For participants affected by COVID-19 infection and/or pandemic while participating in the study, a listing of the following individual COVID-19-related outcome categories will be provided:

- Death due to COVID-19
- Adverse event of COVID-19, as determined by COVID-19 SMQ narrow search
- Specific AE directly associated with the pathogen causing COVID-19, as determined by MST if applicable
- Hospitalization (using data from AE eCRF) due to AE of COVID-19 as determined by COVID-19 SMQ narrow search
- Study drug discontinuation due to COVID-19
- Study discontinuation due to COVID-19
- Missed visits due to COVID-19
- Missed key assessments (contributing to primary and secondary endpoints) due to COVID-19

In addition, composite broad COVID-19 impact indicator will be derived based on the following individual categories defined above: death, AE, hospitalization, study drug discontinuation, study discontinuation, missed visits, and missed key assessments. Composite specific COVID-19 impact indicator will be derived based on death and specific AE.

5. BASELINE CHARACTERISTICS

5.1. Demographics and Baseline Characteristics

Participant demographic variables (ie, age, sex, race, and ethnicity) and baseline characteristics (body weight [in kg], height [in cm], body mass index [BMI; in kg/m^2]) will be summarized by Analysis Group and overall using descriptive statistics for continuous variables and using number and percentage of participants for categorical variables. The summary of demographic data will be provided for the Blinded Study Phase using the Safety Analysis Set and for the OLE Phase using the OLE Analysis Set.

A by-participant demographic listing, including the informed consent date from the last screen (if the participant was rescreened), will be provided by participant ID number in ascending order.

5.2. Other Baseline Characteristics

Other baseline characteristics include:

- ALP and ALP categories ($\leq$ ULN, > 1 to 1.5 × ULN, > 1.5 to 2 × ULN, > 2 to 3 × ULN, or > 3 × ULN)
- GGT and GGT category ($\leq$ ULN or > ULN)
- ALT and ALT category ($\leq$ ULN or > ULN)
- AST and AST category ($\leq$ ULN or > ULN)
- Total bilirubin and total bilirubin category ($\leq$ 1 mg/dL or > 1 mg/dL)
- Direct bilirubin
- Fasting glucose
- Fasting insulin
- HbA_{1c}
- Fasting bile acids
- Fasting triglycerides
- Total cholesterol
- High-density lipoprotein-cholesterol (HDL-C)

- Calculated fasting low-density lipoprotein-cholesterol (LDL-C)
- Calculated fasting very low-density lipoprotein-cholesterol (VLDL-C)
- Calculated fasting nonhigh-density lipoprotein-cholesterol (Non-HDL-C)
- International normalized ratio (INR)
- CRP
- Albumin
- Platelets
- Creatinine
- MELD score
- Child-Pugh (CP) score (summarized by number and percentage)
- Fibrosis stage score according to Ludwig classification (summarized by number and percentage)
- Hepatic collagen content
- ELF test score and components and ELF score category (< 9.8 or ≥ 9.8)
- FibroSURE/FibroTest® and selected components
- FibroScan
- Pruritus VAS score
- 5D-itch score
- UDCA use (yes or no, based on concomitant medication page)
- IBD history (yes or no, based on medical history data)
- Ulcerative colitis (UC) history (yes or no, based on medical history data)
- Crohn's disease history (yes or no, based on medical history data)
- Baseline Partial Mayo Score for patients with IBD history
- Baseline classification of biliary tree disease due to PSC based on MRCP

These baseline characteristics in the Blinded Study Phase will be summarized by Analysis Groups 1 and 2 (Section 3.2) and overall using descriptive statistics for continuous variables and using number and percentage of participants for categorical variables. The summary of these baseline characteristics will be provided for the Safety Analysis Set. No formal statistical testing is planned.

The tables will be repeated for the OLE Phase by Analysis Groups 3 and 4 (Section 3.2), except for biopsy data, fasting insulin, HbA_{1c}, and classification of biliary tree disease due to PSC based on MRCP.

A by-participant listing of other baseline characteristics will be provided by participant ID number in ascending order.

5.3. Medical History

General medical history data will be collected at screening and will be coded using the current version of Medical Dictionary for Regulatory Activities (MedDRA). System organ class (SOC), high-level group term (HLGT), high-level term (HLT), preferred term (PT), and lower-level term (LLT) will be provided in the medical history dataset.

6. EFFICACY ANALYSES

Efficacy data will be summarized and analyzed by Analysis Group (Section 3.2) for the Blinded Study Phase based on the FAS, and for the OLE Phase based on the OLE Analysis Set.

For continuous efficacy endpoints

Descriptive statistics will be provided by Analysis Group as follows:

- Baseline value
- Value at postbaseline visits
- Change from baseline at postbaseline visits
- Percent change from baseline at postbaseline visits (except for QoL, INR, and Mayo risk score [MRS])

Median (Q1, Q3) of the observed values, change from baseline, and percent change from baseline will be plotted using a line plot by Analysis Group and visit for both reporting phases.

For categorical efficacy endpoints

Baseline and postbaseline values will be summarized by counts of participants and percentages at each category level by Analysis Group.

Listings of all efficacy endpoints will be provided for the All Randomized Analysis Set.

6.1. Primary Efficacy Endpoint

6.1.1. Definition of Primary Efficacy Endpoint

The primary efficacy endpoint is the proportion of participants with progression of liver fibrosis, as defined by a $\geq$ 1-stage increase in fibrosis according to the Ludwig classification score at Blinded Study Phase Week 96.

6.1.2. Estimands for the Primary Efficacy Endpoint

The following hybrid estimand of composite and treatment policy strategies (Table 6-1) will be used as the primary estimand for the primary endpoint. Treatment policy strategy will be used to handle all intercurrent events except all-cause mortality and any liver-related clinical events that occur prior to the Week 96 liver biopsy, which will be handled by a composite strategy.

Table 6-1. Estimand Strategy for Primary Endpoint

Primary Endpoints	Hybrid Composite/Treatment Policy Strategy
Population	All participants in the FAS as defined in Section 3.1.2
Patient Level Outcomes to be Measured	≥ 1-stage increase in fibrosis according to the Ludwig classification score at Blinded Study Phase Week 96
Measure of intervention effect and handling of intercurrent events	• Composite policy estimand: participants who experienced a liver-related clinical event[a] or died due to any cause (all-cause mortality) prior to the Week 96 liver biopsy are assumed to be treatment failure (fibrosis progression) for the primary endpoint • Treatment policy estimand: The value of outcome measure is used regardless of the occurrence of intercurrent events other than death or a liver-related clinical event[a] prior to the Week 96 liver biopsy (ie, dose interruption/reduction, protocol deviation, or premature study drug discontinuation, etc)
Population level summary measure	Difference in proportions of fibrosis progression between CILO and placebo
Main Estimators	A stratified Mantel-Haenszel test will be used to compare the difference in proportions of participants who meet the primary endpoint at Week 96 between CILO and placebo, adjusting for baseline UDCA use and fibrosis stage (Ludwig fibrosis score, F3 versus F0-F2) on screening liver biopsy. Participants who experienced a liver-related clinical event[a] or died due to any cause (all-cause mortality) prior to the Week 96 liver biopsy, or had missing data at Week 96 due to reasons other than death or a liver-related clinical event[a] are considered as treatment failure (fibrosis progression).

a Liver-related clinical events include histologic or clinical progression to cirrhosis, hepatic decompensation, liver transplantation, or meeting minimal listing criteria for transplantation (ie, MELD score ≥ 15).

6.1.3. Statistical Hypothesis for Primary Efficacy Endpoint

The statistical hypotheses to be tested can be stated as:

$$H_0: \delta = 0 \quad \text{versus} \quad H_A: \delta < 0,$$

where δ is the difference in the proportion of participants who have progression of liver fibrosis at Week 96 between CILO 100 mg and placebo. The hypothesis will be tested in the FAS with significance level $\alpha = 0.025$. The study will evaluate if the CILO 100 mg group performs better (ie, has a lower proportion of participants with progression of liver fibrosis) than the placebo group.

6.1.4. Primary Analysis of Primary Efficacy Endpoint

A stratified Mantel-Haenszel (MH) test will be used to compare the difference in the proportion of participants who have progression of liver fibrosis at Week 96 between the CILO and placebo groups, adjusting for baseline UDCA use and fibrosis stage (Ludwig fibrosis score, F3 versus F0-F2) on screening liver biopsy. Participants who experienced a liver-related clinical event or died due to any cause (all-cause mortality) prior to the Week 96 liver biopsy, or had missing data at Week 96 due to reasons other than death or a liver-related clinical event will be analyzed as

treatment failures. The point estimate and 2-sided 95% CI for the difference in proportions will be calculated based on stratum-adjusted MH proportions as follows {Koch 1989}:

$$p_A - p_B \pm z_{(1-\alpha)} SE(p_A - p_B),$$

where

- $p_A - p_B = \sum w_s d_s / \sum w_s$ is the stratum-adjusted MH proportion difference, where $d_s = p_{As} - p_{Bs}$ is the difference in the proportion of participants with progression of liver fibrosis between the CILO group and placebo in stratum s.
- $w_s = n_{As} n_{Bs} / (n_{As} + n_{Bs})$, is the weight based on the harmonic mean of sample size per treatment group for each stratum, where n_{As} and n_{Bs} are the sample sizes of the CILO group and placebo, respectively, in stratum s.
- $SE(p_B - p_A) = \sqrt{\sum w_s^2 \left[\frac{p^*_{As}(1-p^*_{As})}{n_{As}-1} + \frac{p^*_{Bs}(1-p^*_{Bs})}{n_{Bs}-1}\right] \Big/ (\sum w_s)^2}$,
  where $p^*_{As} = (m_{As} + 0.5)/(n_{As} + 1)$ and $p^*_{Bs} = (m_{Bs} + 0.5)/(n_{Bs} + 1)$, and m_{As} and m_{Bs} are the number of participants with progression of liver fibrosis in the CILO group and placebo, respectively, in stratum s.
- $z_{(1-\alpha)}$ is the $(1-\alpha)$th percentile of the standard normal distribution
- $\alpha = 0.025$ for the calculation of 95% CI

If the computed lower confidence bound is less than –1, the lower bound is defined as –1. If the computed upper confidence bound is greater than 1, then the upper bound is defined as 1. The point estimates, 95% CIs, and 1-sided p-values based on Z tests, which is $(p_A - p_B)/SE(p_A - p_B)$, will be provided for the differences in proportions.

The primary efficacy objective will be achieved if 1-sided *P*-value of the Z test is < 0.025, or equivalently, if the upper confidence bound of the 95% CI is less than 0.

Forest plots of the treatment differences in proportions will be generated for subgroup analyses as described in Section 3.4.

6.1.5. Sensitivity Analysis of Primary Efficacy Endpoint

The primary analysis as described in Section 6.1.2 to 6.1.4 will be performed after applying each of the following missing data imputation methods: OC, LOCF, MI, and TPA (see details in Section 3.6.1).

6.2. Secondary Efficacy Endpoints

6.2.1. Definition of Secondary Efficacy Endpoints

The secondary endpoints of this study are as follows:

- Changes from baseline in serum concentrations of ALP, ALT, and bile acids at Blinded Study Phase Week 96
- The proportion of participants with ≥ 25% relative reduction in serum ALP concentration from baseline (biochemical response) and no worsening of fibrosis according to the Ludwig classification (histologic response) at Blinded Study Phase Week 96
- The proportion of participants with fibrosis improvement (according to the Ludwig classification) at Blinded Study Phase Week 96
- Changes from baseline in noninvasive markers of fibrosis, including liver stiffness by FibroScan and ELF test score at Blinded Study Phase Week 96
- Change from baseline in PSC Symptoms – Module 1 based on the disease-specific PSC-PRO (Appendix 3) at Blinded Study Phase Week 96

6.2.2. Estimands for the Secondary Efficacy Endpoints

For binary endpoints, the same hybrid estimand strategy as described for the primary endpoints in Section 6.1.2 will be considered. Table 6-2 describes the estimand strategy for the composite endpoint of ALP reduction and no worsening of fibrosis at Blinded Study Phase Week 96.

Table 6-2. Estimand Strategy for the Secondary Endpoint of ALP Reduction and No Worsening of Fibrosis at Blinded Study Phase Week 96

Primary Endpoints	Hybrid Composite/Treatment Policy Strategy
Population	All participants in the FAS as defined in Section 3.1.2
Patient Level Outcomes to be Measured	≥ 25% relative reduction in serum ALP and no increase in fibrosis stage according to the Ludwig classification score at Blinded Study Phase Week 96
Measure of intervention effect and handling of intercurrent events	• Composite policy estimand: participants who experienced a liver-related clinical event[a] or died due to any cause (all-cause mortality) prior to the Week 96 liver biopsy are assumed to be nonresponders • Treatment policy estimand: The value of outcome measure is used regardless of the occurrence of intercurrent events other than death or a liver-related clinical event[a] prior to the Week 96 liver biopsy (ie, dose interruption/reduction, protocol deviation, or premature study drug discontinuation, etc)
Population level summary measure	Difference in proportions of responders between CILO and placebo
Main Estimators	A stratified Mantel-Haenszel test will be used to compare the difference in proportions of participants who meet ≥ 25% relative reduction in serum ALP and no increase in fibrosis stage according to the Ludwig classification at Week 96 between CILO and placebo, adjusting for baseline UDCA use and fibrosis stage (Ludwig fibrosis score, F3 versus F0-F2) on screening liver biopsy. Participants who experienced a liver-related clinical event[a] or died due to any cause (all-cause mortality) prior to the Week 96 liver biopsy, or had missing data at Week 96 due to reasons other than death or a liver-related clinical event[a] are considered nonresponders.

a Liver-related clinical events include histologic or clinical progression to cirrhosis, hepatic decompensation, liver transplantation, or meeting minimal listing criteria for transplantation (ie, MELD score ≥ 15).

Table 6-3 describes the estimand strategy for the secondary endpoint of fibrosis improvement at Blinded Study Phase Week 96.

Table 6-3. Estimand Strategy for the Secondary Endpoint of Fibrosis Improvement at Blinded Study Phase Week 96

Primary Endpoints	Hybrid Composite/Treatment Policy Strategy
Population	All participants in the FAS as defined in Section 3.1.2
Patient Level Outcomes to be Measured	≥ 1-stage decrease in the Ludwig classification score at Blinded Study Phase Week 96
Measure of intervention effect and handling of intercurrent events	• Composite policy estimand: participants who experienced a liver-related clinical event[a] or died due to any cause (all-cause mortality) prior to the Week 96 liver biopsy are assumed to be nonresponders • Treatment policy estimand: The value of outcome measure is used regardless of the occurrence of intercurrent events other than death or a liver-related clinical event[a] prior to the Week 96 liver biopsy (ie, dose interruption/reduction, protocol deviation, or premature study drug discontinuation, etc.)
Population level summary measure	Difference in proportions of responders between CILO and placebo
Main Estimators	A stratified Mantel-Haenszel test will be used to compare the difference in proportions of participants who meet fibrosis improvement at Week 96 between CILO and placebo, adjusting for baseline UDCA use and fibrosis stage (Ludwig fibrosis score, F3 versus F0-F2) on screening liver biopsy. Participants who experienced a liver-related clinical event[a] or died due to any cause (all-cause mortality) prior to the Week 96 liver biopsy, or had missing data at Week 96 due to reasons other than death or a liver-related clinical event[a] are considered nonresponders.

a Liver-related clinical events include histologic or clinical progression to cirrhosis, hepatic decompensation, liver transplantation, or meeting minimal listing criteria for transplantation (ie, MELD score ≥ 15).

For continuous endpoints, the estimand with treatment policy strategy as described in Table 6-4 will be considered.

Table 6-4. Estimand Strategy for the Continuous Secondary Endpoints at Blinded Study Phase Week 96

Primary Endpoints	Treatment Policy Strategy
Population	All participants in the FAS as defined in Section 3.1.2
Patient Level Outcomes to be Measured	Six continuous outcome measures: Change from baseline in serum ALP at Blinded Study Phase Week 96 Change from baseline in serum ALT at Blinded Study Phase Week 96 Change from baseline in serum bile acids at Blinded Study Phase Week 96 Change from baseline in in PSC Symptoms – Module 1 on the disease specific PSC PRO at Blinded Study Phase Week 96 Change from baseline in ELF test score at Blinded Study Phase Week 96 Change from baseline in liver stiffness by FibroScan at Blinded Study Phase Week 96
Measure of intervention effect and handling of intercurrent events	A treatment policy estimand will be considered where the value of outcome measure is used regardless of the occurrence of intercurrent events
Population level summary measure	Mean difference between CILO and placebo for each of the 6 continuous endpoints
Main Estimators	An analysis of covariance (ANCOVA) model will be used to evaluate treatment effect between CILO and placebo for each continuous endpoint at Week 96, adjusting for baseline value of the corresponding endpoint, baseline UDCA use and fibrosis stage (Ludwig fibrosis score, F3 versus F0-F2) on screening liver biopsy. Participants with missing data at Week 96 will be imputed using multiple imputation method as described in Section 3.6.1.

6.2.3. Analysis of the Secondary Efficacy Endpoints

The secondary efficacy endpoints will be tested sequentially in the following order at the same 1-sided significance level of 0.025 after the primary efficacy objective has been achieved. For continuous outcomes, the MI method will be used for imputing missing data and an analysis of covariance (ANCOVA) model will be used for evaluating treatment effect, adjusting for baseline value of the dependent variable, baseline UDCA use, and fibrosis stage (F3 versus F0-F2) on screening liver biopsy. For binary outcomes, the NRI method will be applied for imputing missing values and a stratified Mantel-Haenszel test will be used for evaluating treatment effect, adjusting for baseline UDCA use and fibrosis stage (F3 versus F0-F2) on screening liver biopsy. If a 1-sided P-value ≤ 0.025 is achieved for the corresponding endpoint, the next endpoint will be evaluated; otherwise, testing of the remaining endpoints will cease.

1. Change from baseline in serum ALP at Blinded Study Phase Week 96
2. Change from baseline in serum ALT at Blinded Study Phase Week 96
3. Change from baseline in serum bile acids at Blinded Study Phase Week 96

4. The proportion of participants with ≥ 25% relative reduction in serum ALP concentration from baseline (biochemical response) and no worsening of fibrosis according to the Ludwig classification (histologic response) at Blinded Study Phase Week 96
5. The proportion of participants with fibrosis improvement (according to the Ludwig classification) at Blinded Study Phase Week 96
6. Change from baseline in PSC Symptoms – Module 1 on the disease-specific PSC-PRO at Blinded Study Phase Week 96
7. Change from baseline in ELF test score at Blinded Study Phase Week 96
8. Change from baseline in liver stiffness by FibroScan at Blinded Study Phase Week 96

The point estimates and 2-sided 95% CIs for the differences in least-squares (LS) means (for continuous endpoints) based on ANCOVA model and for the differences in proportions (for binary endpoints) based on stratum-adjusted MH proportions, as well as the corresponding 1-sided p-values will be provided.

6.2.4. Sensitivity Analysis of the Secondary Efficacy Endpoints

LOCF and OC methods (see details in Section 3.6.1) will be carried out to impute missing data for the secondary efficacy endpoints using the same analysis method as described in Sections 6.2.2 and 6.2.3.

CCI [redacted]

CCI [redacted]

CCI [redacted]:

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted] on

- Changes from baseline in the Mayo risk score and Amsterdam-Oxford score
- Incidence of PSC-related complications including ascending cholangitis, dominant strictures, progression to cirrhosis, hepatic decompensation, cholangiocarcinoma, HCC, liver transplantation or meeting minimal listing criteria for transplantation (ie, MELD score ≥ 15), and mortality
- Event-free survival (EFS), defined as time to the first clinical event including histologic or clinical progression to cirrhosis, hepatic decompensation, liver transplantation or meeting minimal listing criteria for transplantation (ie, MELD score ≥ 15), all-cause mortality, or the last follow-up, whichever occurs first

CCI [redacted] **nts**

CCI [redacted]

For survival endpoint: EFS

EFS will be analyzed for Blinded Study Phase only using FAS. The last follow-up date is defined as the latest date among the study drug end date, clinical visit date, laboratory sample collection date, vital signs assessment date, image (Fibroscan and MRCP) assessment date, colonoscopy date, withdrawal of consent date, AE date, concomitant medication date, drug dispensation date and bottle return date for participants who have permanently discontinued study in Blinded Study Phase, or the day prior to the first dosing date of OLE Phase for participants who entered OLE Phase, or the database snapshot date (or data cut date if applicable) for participants who were still on treatment in Blinded Study Phase at the time of the interim analysis.

The Kaplan-Meier (KM) estimates of EFS probability at 3-month intervals, ie, at Months 3, 6, 9, etc and 95% CIs will be provided for each treatment group for the Blinded Study Phase. The 95% CI will be calculated based on Greenwood's formula and log-log transformation of the survival function.

Median, Q1, and Q3 of the EFS will be provided for each treatment group. The 95% CI for median EFS based on Brookmeyer-Crowley method {Brookmeyer 1982} of inverting a generalization of the sign test for censored data will also be provided for each treatment group.

KM curves will be plotted for the clinical EFS by treatment group. A by-participant listing of clinical events will be provided.

No missing data imputation method will be applied to the survival endpoint.

6.4. Changes from Protocol-Specified Efficacy Analyses

There are no deviations from the protocol-specified efficacy analyses.

7. SAFETY ANALYSES

7.1. Adverse Events and Deaths

7.1.1. Adverse Event Dictionary

Clinical and laboratory AEs will be coded using the current version of MedDRA. SOC, HLGT, HLT, PT, and LLT will be provided in the AE dataset.

7.1.2. Adverse Event Severity

Adverse events are graded by the investigator as Grade 1, 2, 3, 4, or 5 according to toxicity criteria specified in the protocol. The severity grade of events for which the investigator did not record severity will be categorized as "missing" for tabular summaries and data listings. The missing category will be listed last in summary presentation.

7.1.3. Relationship of Adverse Events to Study Drug

Related AEs are those for which the investigator selected "Related" on the AE CRF to the question of "Related to Study Treatment." Relatedness will always default to the investigator's choice, not that of the medical monitor. Events for which the investigator did not record relationship to study drug will be considered related to study drug for summary purposes. However, by-participant data listings will show the relationship as missing.

7.1.4. Serious Adverse Events

Serious adverse events (SAEs) will be identified and captured as SAEs if the AEs met the definitions of SAEs that were specified in the study protocol. SAEs captured and stored in the clinical database will be reconciled with the SAE database from the Gilead Global Patient Safety Department before data finalization.

7.1.5. Treatment-Emergent Adverse Events

7.1.5.1. Definition of Treatment-Emergent Adverse Events

For Blinded Study Phase and OLE Phase, treatment-emergent adverse events (TEAEs) are defined as 1 or both of the following:

- Any AEs with an onset date on or after the study drug start date and no later than 30 days after permanent discontinuation of study drug
- Any AEs leading to premature discontinuation of study drug.

For participants entering the OLE Phase, the Blinded Study Phase ends on the day before OLE Day 1.

7.1.5.2. Incomplete Dates

For each reporting phase, if the onset date of the AE is incomplete and the AE stop date is not prior to the first dosing date of study drug, then the month and year (or year alone if month is not recorded) of onset determine whether an AE is treatment emergent. The event is considered treatment emergent if both of the following 2 criteria are met:

- The AE onset is the same as or after the month and year (or year) of the first dosing date of study drug, and
- The AE onset date is the same as or before the month and year (or year) of the date corresponding to 30 days after the date of the last dose of study drug for participants who have permanently discontinued study drug, or the date of database snapshot (or data cut if applicable) for participants who were still on treatment at the time of the interim analysis or primary analysis.

An AE with completely missing onset and stop dates, or with the onset date missing and a stop date later than the first dosing date of study drug, will be considered to be treatment emergent. In addition, an AE with the onset date missing and incomplete stop date with the same or later month and year (or year alone if month is not recorded) as the first dosing date of study drug will be considered treatment emergent.

7.1.6. Summaries of Adverse Events and Deaths

Treatment-emergent AEs will be summarized using number and percentage, and exposure-adjusted incidence rate (EAIR) for selected TEAEs, for the Blinded Study Phase and OLE Phase, where the denominator of EAIR is the total exposure (in 100 person-years) up to the first event of the corresponding category or the last follow-up, whichever occurs earlier, for participants in the Safety Analysis Set and OLE Analysis Set, respectively.

7.1.6.1. Summaries of AE Incidence in Combined Severity Grade Subsets

A brief, high-level summary of the number and percentage and/or EAIR of participants who experienced at least 1 TEAE in the categories described below will be provided by Analysis Group for both reporting phases (Section 3.2). All deaths observed in the study will also be included in this summary.

The number and percentage and/or EAIR of participants who experienced at least 1 TEAE will be provided and summarized by SOC, HLT, PT, and Analysis Group for both reporting phases.

- TEAEs

For the AE categories described below, summaries of the number and percentage will be provided by SOC, PT, and Analysis Group for both reporting phases:

- TEAEs with Grade 2 or higher
- TEAEs with Grade 3 or higher
- Treatment-emergent (TE) treatment-related AEs
- TE treatment-related AEs with Grade 2 or higher
- TE treatment-related AEs with Grade 3 or higher
- TE SAEs
- TE treatment-related SAEs
- TEAEs leading to premature discontinuation of study drug
- TEAEs leading to death (ie, outcome of death)
- TEAEs leading to temporary interruption of study drug
- TEAEs leading to dose reduction of study drug

For TE SAEs and TEAEs leading to premature discontinuation of study drug, summaries of the number and percentage and/or EAIR will be provided for both reporting phases.

Multiple events will be counted only once per participant in each summary. Adverse events will be summarized and listed first in alphabetic order of SOC (and HLT within each SOC if applicable) and then by PT in descending order of frequency in the CILO 100 mg group or Previously on CILO group within each SOC for the Blinded Study Phase and the OLE Phase. For summaries by severity grade, the most severe grade will be used for those AEs that occurred more than once in an individual participant during the study.

In addition to the above summary tables, summaries of number and percentage will be provided for the following tables by PT only, in descending order of frequency in the CILO 100 mg group or Previously on CILO group for the Blinded Study Phase and the OLE Phase:

- TEAEs
- TEAEs of Grade 2 or higher
- TEAEs of Grade 3 or higher
- TE treatment-related AEs

- TE Treatment-related AEs of Grade 2 or higher
- TE Treatment-related AEs of Grade 3 or higher
- TE SAEs
- TE treatment-related SAEs
- TEAEs leading to premature discontinuation of study drug
- TEAEs leading to temporary interruption of study drug
- TEAEs leading to dose reduction of study drug

In addition, data listings will be provided for the following:

- All AEs, indicating whether the event is treatment emergent
- All SAEs
- All Deaths
- All AEs with severity of Grade 3 or higher
- All AEs leading to premature discontinuation of study drug
- All AEs leading to temporary interruption of study drug
- All AEs leading to dose reduction of study drug

7.1.6.2. Summaries of AE Incidence by Severity

For the AE categories described below, summaries will be provided by SOC, PT, and Analysis Group for both reporting phases:

- TEAEs by maximum severity
- TE treatment-related AEs by maximum severity

For TEAEs by maximum severity, summaries of the number and percentage and/or EAIR will be provided for both reporting phases.

Multiple events will be counted only once per participant in each summary. Adverse events will be summarized and listed first in alphabetic order of SOC and then by PT in descending order of frequency in the CILO 100 mg group or Previously on CILO group within each SOC for the Blinded Study Phase and the OLE Phase. For summaries by severity, the most severe severity will be used for those AEs that occurred more than once for a given participant during the study.

7.2. Laboratory Evaluations

Laboratory data collected during the study will be analyzed and summarized using both quantitative and qualitative methods. Summaries of laboratory data will be provided for both reporting phases. For each reporting phase, summaries will include data collected up to the date of last dose of study drug plus 30 days for participants who permanently discontinued study drug, or the date of database snapshot (or data cut if applicable) for participants who were still on treatment at the time of the interim analysis or primary analysis. For participants entering the OLE Phase, the Blinded Study Phase ends on the day before OLE Day 1. Data obtained at the follow-up visit will be included in the by-visit summaries by their nominal visit, regardless of the relevant number of days to the last dose date.

The analysis will be based on values reported in conventional units. When values are below the LOQ, they will be listed as such, and the closest imputed value will be used for the purpose of calculating summary statistics as specified in Section 3.7. Hemolyzed test results will not be included in the analysis, but they will be listed in by-participant laboratory listings.

A by-participant listing for laboratory test results will be provided by participant ID number and visit in chronological order for hematology, serum chemistry, coagulation, and urinalysis separately. Values falling outside of the relevant reference range and/or having a severity grade of 1 or higher on the Common Toxicity Criteria for Adverse Events (CTCAE) severity grade as described in Appendix 2 will be flagged in the data listings, as appropriate.

No formal statistical testing is planned.

7.2.1. Summaries of Numeric Laboratory Results

Descriptive statistics will be provided by Analysis Group for both reporting phases (Section 3.2) for MELD, CRP, HbA_{1c}, creatinine, C-peptide, fasting insulin, fasting glucose, fasting triglycerides, total cholesterol, HDL-C, fasting LDL-C, fasting VLDL-C, fasting non-HDL-C, prothrombin time, white blood count, neutrophils, lymphocytes, hemoglobin, and platelets as follows:

- Baseline values
- Values at each postbaseline visit
- Change from baseline at each postbaseline visit
- Percentage change from baseline at each postbaseline visit (except for MELD)

A baseline laboratory value will be defined as the last measurement obtained on or prior to the date/time of first dose of study drug in a reporting phase. Change from baseline to a postbaseline visit will be defined as the visit value minus the baseline value. The mean, median, Q1, Q3, minimum, and maximum values will be displayed to the reported number of digits; SD values will be displayed to the reported number of digits plus 1.

CP score will be summarized using number and percentage by Analysis Group for both reporting phases at baseline and each postbaseline visit.

In the case of multiple values in an analysis window, data will be selected for analysis as described in Section 3.8.3.

7.2.2. Graded Laboratory Values

The CTCAE Version 5.0 will be used to assign toxicity grades (0 to 4) to laboratory results for analysis. Grade 0 includes all values that do not meet the criteria for an abnormality of at least Grade 1. For laboratory tests with criteria for both increased and decreased levels, analyses for each direction (ie, increased, decreased) will be presented separately.

For ALP, total bilirubin, GGT, ALT, and AST toxicity grades in the Blinded Study Phase and OLE Phase, the CTCAE Version 5.0 will be used to assign grades to the derived baseline values and all postbaseline values.

For all laboratory parameters where the toxicity grades are assigned based on comparing values to the corresponding baseline values, the toxicity grades in the OLE Phase will be reassigned using OLE baseline values according to CTCAE Version 5.0.

7.2.2.1. Treatment-Emergent Laboratory Abnormalities

For Blinded Study Phase and OLE Phase, respectively, treatment-emergent laboratory abnormalities are defined as values that increase at least 1 toxicity grade from baseline at any postbaseline time point, up to and including the date of last dose of study drug plus 30 days for participants who permanently discontinued study drug, or the date of database snapshot (or data cut if applicable) for participants who were still on treatment at the time of the interim analysis or primary analysis. For participants entering the OLE Phase, the Blinded Study Phase ends on the day before OLE Day 1. If the relevant baseline laboratory value is missing, any abnormality of at least Grade 1 observed within the time frame specified above will be considered treatment emergent.

7.2.2.2. Treatment-Emergent Marked Laboratory Abnormalities

For Blinded Study Phase and OLE Phase, respectively, treatment-emergent marked laboratory abnormalities are defined as values that increase from baseline by at least 3 toxicity grades at any postbaseline time point, up to and including the date of the last dose of study drug plus 30 days for participants who permanently discontinued study drug, or the date of database snapshot (or data cut if applicable) for participants who were still on treatment at the time of the interim analysis or primary analysis. For participants entering the OLE Phase, the Blinded Study Phase ends on the day before OLE Day 1. If the relevant baseline laboratory value is missing, any Grade 3 or 4 values observed within the timeframe specified above will be considered treatment-emergent marked abnormalities.

7.2.2.3. Summaries of Laboratory Abnormalities

Laboratory data that are categorical will be summarized using the number and percentage and/or EAIR of participants in the study with the given response at baseline and each scheduled postbaseline visit.

The following summaries (number and percentage and/or EAIR of participants) for treatment-emergent laboratory abnormalities will be provided by lab test and Analysis Group for both reporting phases (Section 3.2). Participants will be categorized according to the most severe postbaseline abnormality grade for a given lab test:

- Graded laboratory abnormalities
- Grade 3 or 4 laboratory abnormalities
- Marked laboratory abnormalities

For all summaries of laboratory abnormalities in the Blinded Study Phase and OLE Phase, the denominator is the number of participants, or total exposure (in 100 person-years) up to the first event or the last follow-up, whichever occurs earlier, for participants in the Safety Analysis Set and OLE Analysis Set, respectively, with at least 1 nonmissing postbaseline value up to 30 days after last dosing date for participants who permanently discontinued study drug, or the date of database snapshot (or data cut if applicable) for participants who were still on treatment at the time of the interim analysis or primary analysis. For participants entering the OLE Phase, the Blinded Study Phase ends on the day before OLE Day 1.

A by-participant listing of treatment-emergent Grade 3 or 4 laboratory abnormalities will be provided by participant ID number and visit in chronological order. This listing will include all test results that were collected throughout the study for the lab test of interest, with all applicable severity grades displayed.

7.2.3. Liver-related Laboratory Evaluations

Liver-related abnormalities after initial study drug dosing will be examined and summarized using the number and percentage and/or EAIR of participants who were reported to have the following laboratory test values for postbaseline measurements by Analysis Group for both reporting phases (Section 3.2).

For participants with normal baseline ALT/AST (≤ ULN):

- Participants meeting criteria for close observation
  - ALT/AST > 3 × ULN
- Participants meeting any 1 of the following criteria for holding study drug
  - ALT/AST > 8 × ULN
  - ALT/AST > 5 × ULN for 2 weeks
  - ALT/AST > 3 × ULN and total bilirubin > 2 × ULN (or direct bilirubin 2 × baseline in participants with Gilbert's syndrome)
  - ALT/AST > 3 × ULN and INR > 1.5 (if not on anticoagulation)

For participants with baseline ALT/AST from > 1 to < 5 × ULN

- Participants meeting criteria for close observation
  - ALT/AST > 2 × Baseline or > 300 U/L
- Participants meeting any 1 of the following criteria for holding study drug
  - ALT/AST > 8 × Baseline or > 500 U/L
  - ALT/AST > 3 × Baseline or > 300 U/L, and total bilirubin > 2 × ULN (or direct bilirubin 2 x Baseline in participants with Gilbert's syndrome)
  - ALT/AST > 3 × Baseline and INR > 1.5 (if not on anticoagulation)

For participants with baseline ALT/AST ≥ 5 × ULN

- Participants meeting criteria for close observation
  - ALT/AST > 2 × Baseline
- Participants meeting criteria for holding study drug
  - ALT/AST > 500 U/L

For the Blinded Study Phase and OLE Phase, the summary will include data from all postbaseline visits up to 30 days after the last dose of study drug for participants who permanently discontinued study drug, or up to the date of database snapshot (or data cut if applicable) for participants who were still on treatment at the time of the interim analysis or primary analysis. For participants entering the OLE Phase, the Blinded Study Phase ends on the day before OLE Day 1. The denominator is the number of participants, or total exposure (in 100 person-years) up to the first event or the last follow-up, whichever occurs earlier, for participants in the Safety Analysis Set for the Blinded Study Phase and the OLE Analysis Set for the OLE Phase who have nonmissing postbaseline values of all relevant tests at the same postbaseline visit date. For individual laboratory tests, participants will be counted once based on the most severe postbaseline values. For the composite endpoint, participants will be counted once when the criteria are met at the same postbaseline visit date. A listing of participants who met at least 1 of the above criteria will be provided.

7.2.4. Drug-Induced Liver Injury (DILI)

Due to the challenge of recognizing and diagnosing DILI in participants with pre-existing hepatic dysfunction, a DILI Adjudication Committee will review potential cases of DILI identified based on laboratory parameters listed in Section 7.2.3. Proportion and/or EAIR of participants with DILI events determined by the Adjudication Committee, including confirmed DILI events (or worsening of hepatic function attributable to study drug could not be excluded) and insufficient data, will be summarized by Analysis Group for both reporting phases (Section 3.2). Multiple events will be counted only once per participant. All adjudicated DILI events will be listed for All Randomized Analysis set.

7.3. Body Weight and Vital Signs

Descriptive statistics will be provided by Analysis Group for both reporting phases (Section 3.2) for body weight and vital signs as follows:

- Baseline value
- Values at each postbaseline visit
- Change from baseline at each postbaseline visit
- Percent change from baseline at each postbaseline visit (only for body weight)

A baseline value is defined as the last available value collected on or prior to the date/time of first dose of study drug in a reporting phase. Change from baseline to a postbaseline visit is defined as the postbaseline value minus the baseline value. Body weight and vital signs measured at unscheduled visits will be included for the baseline value selection.

In the case of multiple values in an analysis window, data will be selected for analysis as described in Section 3.8.3. No formal statistical testing is planned.

A by-participant listing of vital signs will be provided by participant ID number and visit in chronological order. Body weight and BMI will be included in the vital signs listing, if space permits. If not, they will be provided separately.

7.4. Prior and Concomitant Medications

Medications collected at screening and during the study will be coded using the current version of the World Health Organization (WHO) Drug dictionary.

7.4.1. Prior Medications

Prior medications are defined as any medications taken before a participant took the first study drug.

Prior medications will be summarized by preferred name using the number and percentage of participants for each treatment group and overall. A participant reporting the same medication more than once will be counted only once when calculating the number and percentage of participants who received that medication. The summary will be ordered by preferred term in order of descending overall frequency. For drugs with the same frequency, sorting will be done alphabetically.

For the purposes of analysis, any medication with a start date prior to the first dosing date of study drug will be included in the prior medication summary regardless of when the stop date is. If a partial start date is entered the medication will be considered prior unless the month and year (if day is missing) or year (if day and month are missing) of the start date are after the first dosing date. Medications with a completely missing start date will be included in the prior medication summary, unless otherwise specified.

Summaries will be based on the Safety Analysis Set. No formal statistical testing is planned.

7.4.2. Concomitant Medications

Concomitant medications are defined as medications taken while a participant took study drug. Use of concomitant medications will be summarized by preferred name using the number and percentage of participants by Analysis Group for the Blinded Study Phase and OLE Phase (Section 3.2). A participant reporting the same medication more than once will be counted only once when calculating the number and percentage of participants who received that medication. The summary will be ordered by preferred term in descending frequency in the CILO 100 mg group or Previously on CILO group for the Blinded Study Phase and the OLE Phase. For drugs with the same frequency, sorting will be done alphabetically.

For the purposes of analysis, any medications with a start date prior to or on the first dosing date of study drug and continued to be taken after the first dosing date, or started after the first dosing date but prior to or on the last dosing date of study drug will be considered concomitant medications. Medications started and stopped on the same day as the first dosing date or the last dosing date of study drug will also be considered concomitant. Medications with a stop date prior to the date of first dosing date of study drug or a start date after the last dosing date of study drug will be excluded from the concomitant medication summary. If a partial stop date is entered, any medication with the month and year (if day is missing) or year (if day and month are missing) prior to the date of first study drug administration will be excluded from the concomitant medication summary. If a partial start date is entered, any medication with the month and year (if day is missing) or year (if day and month are missing) after the study drug stop date will be excluded from the concomitant medication summary. Medications with completely missing start and stop dates will be included in the concomitant medication summary, unless otherwise specified. Summaries will be based on the Safety Analysis Set for the Blinded Study Phase and the OLE Analysis Set for the OLE Phase. No formal statistical testing is planned.

All prior and concomitant medications (other than per-protocol study drugs) will be provided in a by-participant listing sorted by participant ID number and administration date in chronological order.

7.5. Electrocardiogram Results

Electrocardiogram (ECG) data will not be presented in the CSR since ECGs were not assessed in this study other than as part of the screening process for potential new participants.

7.6. Other Safety Measures

7.6.1. Partial Mayo Score

The Partial Mayo Score is a survey score for the assessment of IBD that considers stool frequency, rectal bleeding, and physician's global assessment. At each study visit, 3 subscore components will be reported:

- Stool frequency: a numeric value ranging from 0 to 3 (where 0 = normal number of stools, 1 = 1 to 2 stools more than normal, 2 = 3 to 4 stools more than normal, 3 = 5 or more stools more than normal)
- Rectal bleeding: a numeric value ranging from 0 to 3 (where 0 = no blood seen, 1 = streaks of blood with stool less than half the time, 2 = obvious blood with stool most of the time, 3 = blood alone passes)
- Physician's global assessment: a numeric value ranging from 0 to 3 (where 0 = normal, 1 = mild disease, 2 = moderate disease, 3 = severe disease)

The Partial Mayo Score will be calculated as the sum of the above 3 subscores. If at least 1 of the subscores is missing, the Partial Mayo Score will be considered missing.

The Partial Mayo Score and its 3 components will be reported for participants with IBD history at the study enrollment by Analysis Group for both reporting phases (Section 3.2).

Descriptive statistics will be provided for the Partial Mayo Score by Analysis Group as follows:

- Baseline value
- Values at each postbaseline visit
- Change from baseline at each postbaseline visit

For each reporting phase, a baseline value is defined as the last available value collected on or prior to the date/time of first dose of study drug. Change from baseline to a postbaseline visit is defined as the postbaseline value minus the baseline value. Partial Mayo Scores calculated at unscheduled visits will be included for the baseline value selection. Participants with change from baseline at any postbaseline visit (including unscheduled visit) ≥ 3 points that occurred prior to 30 days after the last dose of study drug for participants who permanently discontinued study drug, or up to the date of database snapshot (or data cut if applicable) for participants who were still on treatment at the time of the interim analysis or primary analysis will be identified. For participants entering the OLE Phase, the Blinded Study Phase ends on the day before OLE Day 1. Number and percent of these participants will be reported. Maximum change from baseline for these participants will be summarized.

In the case of multiple values in an analysis window, data will be selected for analysis as described in Section 3.8.3. No formal statistical testing is planned.

A by-participant listing of available Partial Mayo Scores, its 3 components, and change from baseline will be provided by participant ID number and visit in chronological order. Any participant with a history of IBD who experiences at least a 3-point increase from baseline in the Partial Mayo Score during the course of the study within 30 days of the last dose will be flagged in the listing.

7.6.2. Pruritus

Descriptive statistics will be provided by Analysis Group for both reporting phases (Section 3.2) for pruritus VAS and 5D-itch scores as follows:

- Baseline values
- Values at each postbaseline visit
- Change from baseline at each postbaseline visit

For each reporting phase, a baseline value is defined as the last measurement obtained on or prior to the date/time of first dose of study drug. Change from baseline to a postbaseline visit is defined as the visit value minus the baseline value. The mean, median, Q1, Q3, minimum, and maximum values will be displayed to the reported number of digits; SD values will be displayed to the reported number of digits plus 1.

In the case of multiple values in an analysis window, data will be selected for analysis as described in Section 3.8.3.

7.7. Changes From Protocol-Specified Safety Analyses

There are no deviations from the protocol-specified safety analyses.

8. PHARMACOKINETIC (PK) ANALYSES

8.1. PK Sample Collection

A single PK blood sample will be collected at the Week 4, 12, 36, 60, and 84 visits in the Blinded Study Phase and at OLE Day 1, Week 4, 24, 48, 72, and 96 in the OLE Phase in all participants.

CCI [redacted]

A population PK model will be developed to characterize the PK of CILO and its metabolites (as applicable). CCI [redacted]
Details of the population PK analysis will be provided in a separate population PK analysis plan.

8.2. PK Analyses Related to Intensive PK Sampling

CCI [redacted] Concentrations of CILO and its metabolites (as applicable) in plasma will be determined using validated bioanalytical assays.

8.2.1. Estimation of PK Parameters

PK parameters will be estimated using Phoenix WinNonlin® software using standard noncompartmental methods. The linear/log trapezoidal rule will be used in conjunction with the appropriate noncompartmental model, with input values for dose level, dosing time, plasma concentration, and corresponding real-time values, based on drug dosing times whenever possible.

All predose sample times before time-zero will be converted to 0.

For area under the curve (AUC), samples BLQ of the bioanalytical assays occurring prior to the achievement of the first quantifiable concentration will be assigned a concentration value of 0 to prevent overestimation of the initial AUC. Samples that are BLQ at all other time points will be treated as missing data in WinNonlin. The nominal time point for a key event or dosing interval (τ) may be used to permit direct calculation of AUC over specific time intervals. The appropriateness of this approach will be assessed by the PK scientist on a profile-by-profile basis.

Pharmacokinetic parameters such as AUC_{tau}, λ_z and $t_{1/2}$ are dependent on an accurate estimation of the terminal elimination phase of drug. The appropriateness of calculating these parameters will be evaluated upon inspection of PK data on a profile-by-profile basis by the PK scientist.

8.2.2. PK Parameters

CCI [redacted] The analytes and parameters presented in Table 8-1 will be used to evaluate the PK objectives of the study. The PK parameters to be estimated in this study are listed and defined in the PK Abbreviations section.

Table 8-1. PK Parameters for Each Analyte

Analyte	Parameters
CILO and its metabolites (if applicable)	AUC_{last}, AUC_{tau}, C_{max}, T_{max}, C_{last}, T_{last}, C_{tau}, λz, CLss/F (CILO only), and $t_{1/2}$

Individual participant concentration data and individual participant PK parameters for CILO and its metabolites (as applicable) will be listed and summarized using descriptive statistics by treatment. Summary statistics (n, mean, SD, coefficient of variation [%CV], median, min, max, Q1, and Q3) will be presented for both individual participant concentration data by time point and individual participant PK parameters by treatment. Moreover, the geometric mean, 95% CI, and the mean and SD of the natural log-transformed values will be presented for individual participant PK parameter data.

Individual concentration data listings and summaries will include all participants with concentration data. The sample size for each time point will be based on the number of participants with nonmissing concentration data at that time point. The number of participants with concentration BLQ will be presented for each time point. For summary statistics, BLQ values will be treated as 0 at predose and one-half of the lower LOQ for postdose time points.

Individual PK parameter data listings and summaries will include all participants for whom PK parameter(s) can be derived. The sample size for each PK parameter will be based on the number of participants with nonmissing data for that PK parameter.

The following tables will be provided for each analyte by treatment:

- Individual participant concentration data and summary statistics
- Individual participant plasma PK parameters and summary statistics

The following figures may be provided for each analyte by treatment:

- Mean (± SD) concentration data versus time (on linear and semilogarithmic scales)
- Median (Q1, Q3) concentration data versus time (on linear and semilogarithmic scales)

Individual, mean, and median postdose concentration values that are ≤ LOQ will not be displayed in the figures and remaining points connected.

PK sampling details by participant, including procedures, differences in scheduled and actual draw times, and sample age will be provided in listings.

9. REFERENCES

Brookmeyer R, Crowley J. A confidence interval for the median survival time. Biometrics 1982;38:29-41.

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

Koch GG, Carr GJ, Amara IA, Stokes ME, Uryniak TJ. Categorical Data Analysis. Chapter 13 in Berry, D.A. (ed.). Statistical Methodology in the Pharmaceutical Sciences. New York: Marcel Dekker, Inc., 1989:pp. 414-21.

Ratitch B, O'Kelly M, Tosiello R. Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. Pharm Stat 2013;12 (6):337-47.

Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York, NY: John Wiley & Sons, Inc; 1987.

10. SOFTWARE

SAS® Software Version 9.4. SAS Institute Inc., Cary, NC, USA.

EAST Version 6. Cytel. Cambridge, MA, USA.

11. SAP REVISION

Revision Date (DD MMM YYYY)	Section	Summary of Revision	Reason for Revision

12. APPENDIX

Appendix 1. Study Procedures Table

Table 12-1. Study Procedures Table – Blinded Study Phase: Screening to Week 48

			Treatment Visits (± 3 days)[p]						
Assessments	Screening[a]	Baseline/ Day 1	Week 4	Week 8	Week 12	Telephone Follow-up Week 16	Week 24	Week 36	Week 48 (± 14 days)
Participant Fasting	X[q]	X	X	X	X		X	X	X
Written Informed Consent	X								
Medical History	X								
Review Inclusion/Exclusion Criteria	X	X							
Physical Examination[d]	X	X	X	X	X		X	X	X
Assess Ascites and HE	X	X	X	X	X		X	X	X
CP and MELD Scores	X	X	X	X	X		X	X	X
Vital Signs[e] and Body Weight	X	X	X	X	X		X	X	X
Height	X								
Partial Mayo Score[b]	X	X	X	X	X		X	X	X
Pruritus VAS and 5D-Itch		X	X	X	X		X	X	X
Health Resource Utilization Questionnaire		X					X		X
Quality of Life Questionnaires[c]		X					X		X
12-lead ECG	X								
Review for Active IBD[n]			X	X	X		X	X	X
FibroScan[i] (if available)	X						X		X
MRCP[j]		X[p]							X
Liver Biopsy[k]	X								
Concomitant Medications	X	X	X	X	X	X	X	X	X

Assessments	Screening[a]	Baseline/ Day 1	Treatment Visits (± 3 days)[p] Week 4	Week 8	Week 12	Telephone Follow-up Week 16	Week 24	Week 36	Week 48 (± 14 days)
Adverse Events	X	X	X	X	X	X	X	X	X
Dispense Study Drugs		X	X	X	X		X	X	X
Review Study Drug Compliance			X	X	X	X	X	X	X
Laboratory Assessments									
Chemistry, eGFR, Hematology, Coagulation Panel	X	X	X	X	X		X	X	X
Lipid Profile		X			X		X	X	X
C-Peptide, Insulin and HbA1c		X			X		X		X
HIV-1, HBV and HCV Serology	X								
Urine Drug Screen	X								
Pregnancy Testing[r]	X	X	X	X	X		X	X	X
Serum FSH[s]	X								
Stool Collection (if available)		X							X
Blood for Biomarkers	X	X	X		X		X		X
Single PK and PD Sampling			X		X			X	
CCI [redacted]									

Study Procedures Table – Blinded Study Phase: Week 60 to Follow-Up

	Treatment Visits (±3 days)[e]					
Assessments	Week 60	Week 72	Week 84	Week 96 (± 14 days)	ET[a]	Follow-Up Visit (± 5 days)[r]
Participant Fasting	X	X	X	X	X	X
Physical Examination	X	X	X	X	X	X
Assess Ascites and HE	X	X	X	X	X	X
CP and MELD Scores	X	X	X	X	X	
Vital Signs[e] and Body Weight	X	X	X	X	X	X
Partial Mayo Score[b]	X	X	X	X	X	X
Pruritus VAS and 5D-Itch	X	X	X	X	X	X
Health Resource Utilization Questionnaire		X		X	X	X
Quality of Life Questionnaires[c]		X		X	X	X
Review for Active IBD[m]	X	X	X	X	X	X
FibroScan[i] (if available)		X		X	X	
MRCP[j]				X	X	
Liver Biopsy[k]				X	X	
Concomitant Medications	X	X	X	X	X	X
Adverse Events	X	X	X	X	X	X
Dispense Study Drugs	X	X	X			
Review Study Drug Compliance	X	X	X	X	X	
Laboratory Assessments						
Chemistry, eGFR, Hematology, Coagulation Panel	X	X	X	X	X	X
Lipid Profile	X	X	X	X	X	X
C-Peptide, Insulin, and HbA1c				X	X	
Pregnancy Testing[f]	X	X	X	X	X	X
Stool Collection (if available)				X		

Assessments	Treatment Visits (±3 days)[o]					
	Week 60	Week 72	Week 84	Week 96 (± 14 days)	ET[n]	Follow-Up Visit (± 5 days)[r]
Blood for Biomarkers		X		X	X	X
Single PK and PD Sampling	X		X		X	
CCI						
CCI						

CP = Child-Pugh; ECG = electrocardiogram; eGFR = estimated glomerular filtration rate; ET = early termination; FSH = follicle-stimulating hormone; HbA1c = hemoglobin A1c; HE = hepatic encephalopathy; IBD = inflammatory bowel disease; MELD = Model for End-stage Liver Disease; MRCP = Magnetic Resonance Cholangiopancreatography; PD = pharmacodynamic(s); PK = pharmacokinetic(s); VAS = visual analogue scale

a Participants will be screened within 10 weeks before randomization. The screening period may be extended under special circumstances with the explicit approval of the medical monitor.

b Partial Mayo score calculation for participants with a history of IBD (as appropriate). See Section 6.9.6 of the protocol for details.

c Quality of life (QoL) questionnaires to include: Short Inflammatory Bowel Disease Questionnaire (for participants with a history of IBD), Chronic Liver Disease Questionnaire, EuroQol (5 dimensions [EQ-5D]), and primary sclerosing cholangitis-patient-reported outcome. It is recommended that Health Resource Utilization and QoL questionnaires are completed prior to any study procedures being performed and prior to the participant seeing a health care provider.

d Symptom-driven physical examination. The focus of a symptom-driven physical examination will be determined by the investigator based on participant complaint. A complete physical examination will be completed at screening.

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

f Females of childbearing potential only (see Appendix 4 of the protocol). Serum pregnancy test at screening and urine pregnancy tests at all other visits except telephone follow-up visits.

g Women of any age with amenorrhea of ≥ 12 months (see Appendix 4 of the protocol).

h CCI

i Participant should be in fasted state for FibroScan collection. Refer to the Site Operations Manual for further details.

j If the participant has any contraindications to magnetic resonance imaging, the MRCP is not required. Refer to the MRCP imaging guidelines manual for additional instructions.

k A historical liver biopsy within 6 months of screening may be used if deemed acceptable by the central reader. For ET visit, perform liver biopsy at the discretion of the investigator.

l CCI

m For participants with history of IBD, any evidence of active IBD seen on routinely performed colonoscopy will be captured.

n Participants prematurely discontinuing should complete an ET visit (see Section 6.4.6 of the protocol).

o Treatment visit windows are ± 3 days unless otherwise stated (see Section 6.4 of the protocol).

p MRCP visit window for baseline/Day 1 is + 14 days. Historical MRCP within 3 months of screening visit or a routinely performed MRCP within screening period may be used.

q Fasting for FibroScan per standard of care

r If the Blinded Study Phase follow-up visit and the OLE Phase baseline/Day 1 visit occur on the same day, study assessment procedures should be performed according to the OLE Phase baseline/Day 1

Table 12-2. Study Procedures Table – Open-Label Extension Phase: Baseline/Day 1 to Follow-Up

Assessments	Baseline/ Day 1	Treatment Visits (± 5 days)[i]								
		Week 4	Telephone Follow-up Week 8	Telephone Follow-up Week 12	Week 24	Week 48	Week 72	Week 96 (± 14 days)	ET[h]	Follow- Up Visit (± 7 days)
Participant Fasting	X[j]	X			X	X	X	X	X	X
Written Informed Consent	X[k]									
Physical Examination[a]	X	X			X	X	X	X	X	X
Assess Ascites and HE	X	X			X	X	X	X	X	X
CP and MELD Scores	X	X			X	X	X	X	X	
Vital Signs[b] and Body Weight	X	X			X	X	X	X	X	X
Review for Active IBD[g]	X									
Partial Mayo Score[c] (IBD only)	X	X			X	X	X	X	X	X
Pruritus VAS and 5D-Itch	X	X			X	X	X	X	X	X
Health Resource Utilization Questionnaire[d]	X				X	X	X	X	X	X
Quality of Life Questionnaires[d]	X				X	X	X	X	X	X
FibroScan (if available)[e]	X				X	X	X	X	X	
Concomitant Medications	X	X	X	X	X	X	X	X	X	X
Adverse Events	X	X	X	X	X	X	X	X	X	X
Dispense Study Drugs	X	X			X	X	X			
Review Study Drug Compliance		X	X	X	X	X	X	X		
Laboratory Assessments										
Chemistry, eGFR, Hematology, Coagulation Panel	X	X			X	X	X	X	X	X
Lipid Profile	X				X	X	X	X	X	X

Assessments	Baseline/ Day 1	Treatment Visits (± 5 days)[i]								
		Week 4	Telephone Follow-up Week 8	Telephone Follow-up Week 12	Week 24	Week 48	Week 72	Week 96 (± 14 days)	ET[h]	Follow- Up Visit (± 7 days)
Pregnancy Testing[f]	X	X			X	X	X	X	X	X
Blood for Biomarkers	X					X		X	X	X
Single PK Sampling	X	X			X	X	X	X	X	

CLDQ = Chronic Liver Disease Questionnaire; CP = Child-Pugh; ECG = electrocardiogram; eGFR = estimated glomerular filtration rate; EQ-5D = EuroQol (5 dimensions); ET = early termination; HE = hepatic encephalopathy; IBD = inflammatory bowel disease; MELD = Model for End-stage Liver Disease; PK = pharmacokinetic(s); PSC-PRO = primary sclerosing cholangitis-patient-reported outcome; SIBDQ = Short Inflammatory Bowel Disease Questionnaire; VAS = visual analogue scale

a Symptom-driven physical examination. The focus of a symptom-driven physical examination will be determined by the investigator based on participant complaint. A complete physical examination to be completed at OLE Phase baseline/ Day 1 visit.
b Vital signs include heart rate, systolic and diastolic blood pressure, respiratory rate, and body temperature.
c Partial Mayo score calculation for participants with a history of IBD (as appropriate). See Section 6.9.6 of the protocol for details.
d Quality of life (QoL) questionnaires include: SIBDQ (for participants with a history of IBD), CLDQ, EQ-5D, and PSC-PRO. It is recommended that QoL questionnaires are completed prior to any study procedures being performed and prior to the participant seeing a health care provider.
e Participant should be in fasted state for FibroScan collection. Refer to the Site Operations Manual for further details. FibroScan is not required at OLE Phase baseline/Day 1 if FibroScan was completed within 90 days of the OLE Phase baseline/Day 1 visit.
f Females of childbearing potential only (see Appendix 4 of the protocol). Urine pregnancy test at all visits, except telephone follow-up visits.
g For participants with history of IBD, any evidence of active IBD seen on routinely performed colonoscopy will be captured.
h Participants prematurely discontinuing should complete an OLE Phase ET visit (see Section 6.5.4 of the protocol).
i Treatment visit windows are ± 5 days unless otherwise stated, with the exception of the OLE Phase baseline/Day 1 visit (see Section 6.4 of the protocol).
j Fasting for FibroScan per standard of care.
k Re-consent at OLE Phase baseline/Day 1 visit is not necessary if the participant has previously provided written informed consent to participate in the OLE Phase and there has been no update to the informed consent form since participant previously provided written informed consent.

Appendix 2. CTCAE Grade for Laboratory Parameters

CTCAE 5.0	CTCAE Grade				
CTCAE Term	**Grade 1**	**Grade 2**	**Grade 3**	**Grade 4**	**Grade 5**
Anemia	Hemoglobin (Hgb) <LLN - 10.0 g/dL; <LLN - 6.2 mmol/L; <LLN - 100 g/L	Hgb <10.0 - 8.0 g/dL; <6.2 - 4.9 mmol/L; <100 - 80g/L	Hgb <8.0 g/dL; <4.9 mmol/L; <80 g/L; transfusion indicated	Life-threatening consequences; urgent intervention indicated	Death
Activated partial thromboplastin time prolonged	>ULN - 1.5 x ULN	>1.5 - 2.5 x ULN	>2.5 x ULN; bleeding	-	-
Alanine aminotransferase increased	>ULN - 3.0 x ULN if baseline was normal; 1.5 - 3.0 x baseline if baseline was abnormal	>3.0 - 5.0 x ULN if baseline was normal; >3.0 - 5.0 x baseline if baseline was abnormal	>5.0 - 20.0 x ULN if baseline was normal; >5.0 - 20.0 x baseline if baseline was abnormal	>20.0 x ULN if baseline was normal; >20.0 x baseline if baseline was abnormal	-
Alkaline phosphatase increased	>ULN - 2.5 x ULN if baseline was normal; 2.0 - 2.5 x baseline if baseline was abnormal	>2.5 - 5.0 x ULN if baseline was normal; >2.5 - 5.0 x baseline if baseline was abnormal	>5.0 - 20.0 x ULN if baseline was normal; >5.0 - 20.0 x baseline if baseline was abnormal	>20.0 x ULN if baseline was normal; >20.0 x baseline if baseline was abnormal	-
Aspartate aminotransferase increased	>ULN - 3.0 x ULN if baseline was normal; 1.5 - 3.0 x baseline if baseline was abnormal	>3.0 - 5.0 x ULN if baseline was normal; >3.0 - 5.0 x baseline if baseline was abnormal	>5.0 - 20.0 x ULN if baseline was normal; >5.0 - 20.0 x baseline if baseline was abnormal	>20.0 x ULN if baseline was normal; >20.0 x baseline if baseline was abnormal	-
Blood bicarbonate decreased	<LLN and no intervention initiated	-	-	-	-
Blood bilirubin increased	>ULN - 1.5 x ULN if baseline was normal; > 1.0 - 1.5 x baseline if baseline was abnormal	>1.5 - 3.0 x ULN if baseline was normal; >1.5 - 3.0 x baseline if baseline was abnormal	>3.0 - 10.0 x ULN if baseline was normal; >3.0 - 10.0 x baseline if baseline was abnormal	>10.0 x ULN if baseline was normal; >10.0 x baseline if baseline was abnormal	-
Cholesterol high	>ULN - 300 mg/dL; >ULN - 7.75 mmol/L	>300 - 400 mg/dL; >7.75 - 10.34 mmol/L	>400 - 500 mg/dL; >10.34 - 12.92 mmol/L	>500 mg/dL; >12.92 mmol/L	-
CPK increased	>ULN - 2.5 x ULN	>2.5 x ULN - 5 x ULN	>5 x ULN - 10 x ULN	>10 x ULN	-
Creatinine increased	>ULN - 1.5 x ULN	>1.5 - 3.0 x baseline; >1.5 - 3.0 x ULN	>3.0 x baseline; >3.0 - 6.0 x ULN	>6.0 x ULN	-

CTCAE 5.0	CTCAE Grade				
CTCAE Term	**Grade 1**	**Grade 2**	**Grade 3**	**Grade 4**	**Grade 5**
GGT increased	>ULN - 2.5 x ULN if baseline was normal; 2.0 - 2.5 x baseline if baseline was abnormal	>2.5 - 5.0 x ULN if baseline was normal; >2.5 - 5.0 x baseline if baseline was abnormal	>5.0 - 20.0 x ULN if baseline was normal; >5.0 - 20.0 x baseline if baseline was abnormal	>20.0 x ULN if baseline was normal; >20.0 x baseline if baseline was abnormal	-
Haptoglobin decreased	<LLN	-	-	-	-
Hemoglobin increased	Increase in >0 - 2 g/dL	Increase in >2 - 4 g/dL	Increase in >4 g/dL	-	-
INR increased	>1.2 - 1.5; >1 - 1.5 x baseline if on anticoagulation; monitoring only indicated	>1.5 - 2.5; >1.5 - 2.5 x baseline if on anticoagulation; dose adjustment indicated	>2.5; >2.5 x baseline if on anticoagulation; bleeding	-	-
Lipase increased	>ULN - 1.5 x ULN	>1.5 - 2.0 x ULN; >2.0 - 5.0 x ULN and asymptomatic	>2.0 - 5.0 x ULN with signs or symptoms; >5.0 x ULN and asymptomatic	>5.0 x ULN and with signs or symptoms	-
Lymphocyte count decreased	<LLN - 800/mm3; <LLN - 0.8 x 10e9/L	<800 - 500/mm3; <0.8 - 0.5 x 10e9 /L	<500 - 200/mm3; <0.5 - 0.2 x 10e9 /L	<200/mm3; <0.2 x 10e9 /L	-
Lymphocyte count increased	-	>4000/mm3 - 20,000/mm3	>20,000/mm3	-	-
Neutrophil count decreased	<LLN - 1500/mm3; <LLN - 1.5 x 10e9 /L	<1500 - 1000/mm3; <1.5 - 1.0 x 10e9 /L	<1000 - 500/mm3; <1.0 - 0.5 x 10e9 /L	<500/mm3; <0.5 x 10e9 /L	-
Platelet count decreased	<LLN - 75,000/mm3; <LLN - 75.0 x 10e9 /L	<75,000 - 50,000/mm3; <75.0 - 50.0 x 10e9 /L	<50,000 - 25,000/mm3; <50.0 - 25.0 x 10e9 /L	<25,000/mm3; <25.0 x 10e9 /L	-
Serum amylase increased	>ULN - 1.5 x ULN	>1.5 - 2.0 x ULN; >2.0 - 5.0 x ULN and asymptomatic	>2.0 - 5.0 x ULN with signs or symptoms; >5.0 x ULN and asymptomatic	>5.0 x ULN and with signs or symptoms	-
White blood cell decreased	<LLN - 3000/mm3; <LLN - 3.0 x 10e9 /L	<3000 - 2000/mm3; <3.0 - 2.0 x 10e9 /L	<2000 - 1000/mm3; <2.0 - 1.0 x 10e9 /L	<1000/mm3; <1.0 x 10e9 /L	-
Hypercalcemia	Corrected serum calcium of >ULN - 11.5 mg/dL; >ULN - 2.9 mmol/L; Ionized calcium >ULN - 1.5 mmol/L	Corrected serum calcium of >11.5 - 12.5 mg/dL; >2.9 - 3.1 mmol/L; Ionized calcium >1.5 - 1.6 mmol/L; symptomatic	Corrected serum calcium of >12.5 - 13.5 mg/dL; >3.1 - 3.4 mmol/L; Ionized calcium >1.6 - 1.8 mmol/L; hospitalization indicated	Corrected serum calcium of >13.5 mg/dL; >3.4 mmol/L; Ionized calcium >1.8 mmol/L; life-threatening consequences	Death

CTCAE 5.0	CTCAE Grade				
CTCAE Term	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Hyperkalemia	>ULN - 5.5 mmol/L	>5.5 - 6.0 mmol/L; intervention initiated	>6.0 - 7.0 mmol/L; hospitalization indicated	>7.0 mmol/L; life-threatening consequences	Death
Hypermagnesemia	>ULN - 3.0 mg/dL; >ULN - 1.23 mmol/L	-	>3.0 - 8.0 mg/dL; >1.23 - 3.30 mmol/L	>8.0 mg/dL; >3.30 mmol/L; life-threatening consequences	Death
Hypernatremia	>ULN - 150 mmol/L	>150 - 155 mmol/L; intervention initiated	>155 - 160 mmol/L; hospitalization indicated	>160 mmol/L; life-threatening consequences	Death
Hypertriglyceridemia	150 mg/dL - 300 mg/dL; 1.71 mmol/L - 3.42 mmol/L	>300 mg/dL - 500 mg/dL; >3.42 mmol/L - 5.7 mmol/L	>500 mg/dL - 1000 mg/dL; >5.7 mmol/L - 11.4 mmol/L	>1000 mg/dL; >11.4 mmol/L; life-threatening consequences	Death
Hyperuricemia	>ULN without physiologic consequences	-	>ULN with physiologic consequences	Life-threatening consequences	Death
Hypoalbuminemia	<LLN - 3 g/dL; <LLN - 30 g/L	<3 - 2 g/dL; <30 - 20 g/L	<2 g/dL; <20 g/L	Life-threatening consequences; urgent intervention indicated	Death
Hypocalcemia	Corrected serum calcium of <LLN - 8.0 mg/dL; <LLN - 2.0 mmol/L; Ionized calcium <LLN - 1.0 mmol/L	Corrected serum calcium of <8.0 - 7.0 mg/dL; <2.0 - 1.75 mmol/L; Ionized calcium <1.0 - 0.9 mmol/L; symptomatic	Corrected serum calcium of <7.0 - 6.0 mg/dL; <1.75 - 1.5 mmol/L; Ionized calcium <0.9 - 0.8 mmol/L; hospitalization indicated	Corrected serum calcium of <6.0 mg/dL; <1.5 mmol/L; Ionized calcium <0.8 mmol/L; life-threatening consequences	Death
Hypoglycemia	<LLN - 55 mg/dL; <LLN - 3.0 mmol/L	<55 - 40 mg/dL; <3.0 - 2.2 mmol/L	<40 - 30 mg/dL; <2.2 - 1.7 mmol/L	<30 mg/dL; <1.7 mmol/L; life-threatening consequences; seizures	Death
Hypokalemia	<LLN - 3.0 mmol/L	Symptomatic with <LLN - 3.0 mmol/L; intervention indicated	<3.0 - 2.5 mmol/L; hospitalization indicated	<2.5 mmol/L; life-threatening consequences	Death
Hypomagnesemia	<LLN - 1.2 mg/dL; <LLN - 0.5 mmol/L	<1.2 - 0.9 mg/dL; <0.5 - 0.4 mmol/L	<0.9 - 0.7 mg/dL; <0.4 - 0.3 mmol/L	<0.7 mg/dL; <0.3 mmol/L; life-threatening consequences	Death
Hyponatremia	<LLN - 130 mmol/L	125-129 mmol/L and asymptomatic	125-129 mmol/L symptomatic; 120-124 mmol/L regardless of symptoms	<120 mmol/L; life-threatening consequences	Death

Note: Refer to Common Terminology Criteria for Adverse Events (CTCAE) Version 5.0, which can be found at https://evs nci.nih.gov/ftp1/CTCAE/CTCAE_5.0

Appendix 3. Programming Specifications

A3.1 Enhanced Liver Fibrosis (ELF)

ELF™ test score = 2.278 + 0.851× ln (hyaluronic acid) + 0.751 × ln (P3NP) + 0.394 × ln (TIMP1)

Note: All ELF™ test score components (hyaluronic acid, P3NP and TIMP1) need to be measured from the same blood draw. When ELF™ test score is missing and components are less than the LOQ or above the upper LOQ, the components are to be imputed per data handling conventions in Section 3.7, and ELF™ test score will be calculated based on the imputed values of components.

A3.2 FibroSURE/FibroTest® Calculation

Step-1 formula	f5 = 4.467 × Log[α_2-macroglobulin(g/L)] – 1.357 × Log[Haptoglobin (g/L)] + 1.017 × Log[GGT(U/L)] + 0.0281 × [Age (year)] + 1.737 × Log [Total Bilirubin (umol/L)] – 1.184 × [ApoA1 (g/L)] + 0.301 × Sex (female = 0, male = 1) – 5.540
Step-2 formula	FibroSURE/FibroTest® Score = 1/(1+exp^(-f5))
Note	In the formula, SI value and units should be applied. The Log function in the formula is with base 10. FibroSURE/FibroTest® score should be calculated from the parameters from the same blood draw Age is when the blood draw was taken

For participant with Gilbert's syndrome or other cause of unconjugated hyperbilirubinemia according to the medical history page at screening, the FibroSURE/FibroTest® score will be calculated using Direct Bilirubin instead of Total Bilirubin in above formula throughout the study.

A3.3 Child-Pugh (CP) score

CP score is obtained by adding the score for each parameter from the following table, where the score for each parameter should come from the same day and the lab parameters should come from the same blood draw. If the CP score is missing for all visits between Screening and Baseline/Day 1, then impute the baseline CP score by 1) selecting the latest non-missing HE and Ascites values from the same day on or prior to Baseline/Day 1; 2) selecting non-missing lab parameters from the same blood draw on or prior to Baseline/Day 1 that are closest to the HE and Ascites date (the latter will be selected if there are 2 dates that are equidistant from the HE and Ascites date).

	1	2	3
Hepatic encephalopathy (HE)	**None** No encephalopathy and not on any treatment for hepatic encephalopathy	**Medication-controlled** Participant is lethargic, may have moderate confusion Participant is receiving medical therapy for HE	**Medication-refractory** Marked confusion/incoherent, rousable but sleeping or comatose
Ascites	**None** No ascites and not on treatment for ascites	**Mild/Moderate** Cross sectional imaging showing ascites Abdominal distension Medication for ascites	**Severe (Diuretic-refractory)** Visible clinically
Bilirubin (mg/dL)	< 2	2-3	> 3
Albumin (g/dL)	> 3.5	2.8-3.5	< 2.8
INR	< 1.7	1.7-2.3	> 2.3

A3.4 Model for End-stage Liver Disease (MELD) Score

MELD score = 3.78 [ln total bilirubin (mg/dL)] + 11.2 [ln INR] + 9.57 [ln creatinine (mg/dL)] + 6.43.

Round total bilirubin to 1 decimal place and serum creatinine to 2 decimal places prior to using values in formula or calculation criteria. Lab values less than 1 are set to 1.

Creatinine is set to 4 if greater than 4 or if "For participants on dialysis, did the participants have 2 or more dialysis treatments within the prior week?" is answered as "Yes". Missing answers to dialysis question is imputed as "No". If the creatinine is resulted as "Icteric – Test Not Performed", the calculation will use serum enzymatic creatinine.

Round MELD score to integer. The lab parameters need to be measured from the same blood draw.

A3.5 Mayo Risk Score (MRS)

MRS = 0.03 age (years) + 0.54 Ln total bilirubin (mg/dL) + 0.54 Ln AST (U/L) + 1.24 variceal bleeding (0/1) – 0.84 albumin (g/dL).

Keep 2 decimal places. The lab parameters need to be measured from the same blood draw. Age should be the actual age at the date when lab values are taken. Status of variceal bleeding should also be decided on the lab date. CCI [redacted]

A3.6 Amsterdam-Oxford Score (AOS)

AOS = 0.323 PSC subtype + 0.018 age – 2.485 $\log_{10}$(albumin) + 2.451 abs[$\log_{10}$(Platelets) – 0.5] + 0.347 $\log_{10}$ (AST) + 0.393 $\log_{10}$ (ALP) + 0.337 $\log_{10}$(total bilirubin).

PSC subtype = 1 if large duct PSC, and = 0 if small duct PSC. AST, ALP and total bilirubin are expressed in × ULN. Albumin and platelets are expressed in × lower limit of normal (LLN).

Keep 2 decimal places. The lab parameters need to be measured from the same blood draw. Age should be the actual age at the PSC diagnosis. Incomplete date of birth will be imputed per data handling conventions in Section 3.7. Incomplete date of PSC diagnois will be imputed the same way as incomlete date of birth.

A3.7 Measures of Biliary Stricture Severity by MRCP

The biliary stricture severity as measured by MRCP will be summarized by the last 5 risk scores (#19-23) as described in the following table:

	MRCP Derived Scores	
1.	CBD Stricture Score	• 1, if CBD Stricture Presence = 'N', • 2, if CBD Stricture Presence = 'Y' and CBD Stricture Grade <= 75% • 3, if CBD Stricture Presence = 'Y' and CBD Stricture Grade > 75%
2.	LHD Stricture Score	• 1, if LHD Stricture Presence = 'N' • 2, if LHD Stricture Presence = 'Y' and LHD Stricture Grade <= 75% • 3, if LHD Stricture Presence = 'Y' and LHD Stricture Grade > 75%
3.	RHD Stricture Score	• 1, if RHD Stricture Presence = 'N' • 2, if RHD Stricture Presence = 'Y' and RHD Stricture Grade <= 75% • 3, if RHD Stricture Presence = 'Y' and RHD Stricture Grade > 75%
4.	IHBD Stricture Score	• 1, if IHBD Stricture Presence in Right Lobe = 'N' and IHBD Stricture Presence in Left Lobe = 'N' and IHBD Stricture Presence in Caudate = 'N' • 2, if at least one in IHBD Stricture Presence in Right Lobe, IHBD Stricture Presence in Left Lobe and IHBD Stricture Presence in Caudate = 'Y' and the corresponding grade <= 50% if 'Y' • 3, otherwise (at least one present with grade > 50%)
5.	CBD Stricture Length Score	• 1, if . < CBD Stricture Length <= 2mm or CBD Stricture Presence = 'N' • 2, if 3mm <= CBD Stricture Length <= 10mm • 3, if CBD Stricture Length > 10mm
6.	LHD Stricture Length Score	• 1, if . < LHD Stricture Length <= 2mm or LHD Stricture Presence = 'N' • 2, if 3mm <= LHD Stricture Length <= 10mm • 3, if LHD Stricture Length > 10mm
7.	RHD Stricture Length Score	• 1, if . < RHD Stricture Length <= 2mm or RHD Stricture Presence = 'N' • 2, if 3mm <= RHD Stricture Length <= 10mm • 3, if RHD Stricture Length > 10mm
8.	IHBD Involvement Score	• 0, if IHBD Stricture Presence in Right Lobe = 'N' and IHBD Stricture Presence in Left Lobe = 'N' and IHBD Stricture Presence in Caudate = 'N'; • 1, if one of IHBD Stricture Presence in Right Lobe, IHBD Stricture Presence in Left Lobe and IHBD Stricture Presence in Caudate =' Y' • 2, if two of IHBD Stricture Presence in Right Lobe, IHBD Stricture Presence in Left Lobe and IHBD Stricture Presence in Caudate = 'Y' • 3, if all three of IHBD Stricture Presence in Right Lobe, IHBD Stricture Presence in Left Lobe and IHBD Stricture Presence in Caudate = 'Y'

	MRCP Derived Scores	
9.	CBD Dilatation Score	• 1, if . < CBD Diameter <= 10mm • 2, if 11mm <= CBD Diameter <= 14mm • 3, if CBD Diameter >= 15mm
10.	LHD Dilatation Score	• 1, if .< LH Diameter <= 6mm • 2, if 7mm <= LH Diameter <= 8mm • 3, if LH Diameter >= 9mm
11.	RHD Dilatation Score	• 1, if . < RH Diameter <= 6mm • 2, if 7mm <= RH Diameter <= 8mm • 3, if RH Diameter >= 9mm
12.	IHBD Dilatation Score	• 1, if . < IHB Diameter <= 3mm • 2, if IHB Diameter = 4mm • 3, if IHB Diameter >= 5mm
13.	Parenchymal enhancement heterogeneity Score	• 0, if Heterogeneity = 'N' • 1, if Heterogeneity = 'Y'
14.	Stone Score	• 0, if Stone Presence = 'N' • 1, if Stone Presence = 'Y'
15.	Dysmorphy Score	• 0, if Significant Atrophy='N' and Marked Lobulation='0' and Increased Caudate/Rt Lobe Ratio<=0.6 • 1, otherwise and if Significant Atrophy and Marked Lobulation and Increased Caudate/Rt Lobe Ratio are non-missing
16.	Portal Hypertension Score	• 0, if Portal Hypertension = 'N' • 1, if Portal Hypertension = 'Y'
17.	Gall Bladder Score	• 0, if Gall Bladder Presence = 'N' • 1, if Gall Bladder Presence = 'Y' and Gall Bladder Length/Gall Bladder Width >= 2 • 2, if Gall Bladder Presence = 'Y' and Gall Bladder Length/Gall Bladder Width < 2
18.	Lymph nodes Score	• 0, if Perihepatic Supracentimetric Lymph Nodes = 'N' • 1, if Perihepatic Supracentimetric Lymph Nodes = 'Y'
19.	Total MRCP Score	• Summation of the above18 derived scores
20.	MRI Progression Risk Score	• 1 × IHBD Dilatation Score + 2 × Dysmorphy Score + 1 × Portal Hypertension Score
21.	MRCP Risk Score (PSC Survival)	• 1 × Dysmorphy Score+ 1 × Portal Hypertension + 1 × Lymph nodes Score
22.	MRCP Risk Score (Cholangitis Survival)	• 1 × Portal Hypertension + 1 × Lymph nodes Score
23.	MRCP Risk Score (Non-Cholangitis Survival)	• 1 × Portal Hypertension + 1 × Lymph nodes Score + 0.5 × CBD Stricture Score

A3.8 Health Related QoL Score Calculations

PSC-PRO Module 1 – PSC Symptoms Total Score:

PSC-PRO module 1 – PSC symptoms contains a total of 12 questions asking about the severity of specific PSC symptoms on a scale of 0 (no symptoms) to 10 (symptoms as bad as you could imagine) with a 24-hour recall period. The total score, which is computed as 12 times the average of nonmissing scores of the 12 questions, can potentially range between 0 and 120. For participants with missing responses to 6 or more questions, the total score will be considered as missing.

5D-Itch Score:

5D-Itch contains a total of 5 questions with each ranging from 1 to 5. It includes 16 potential locations of itch, including 15 body part items and one point of contact with clothing or bandages. Single-item domain scores (duration, degree and direction) are equal to the value indicated below the response choice (range 1–5). The score for the disability domain is achieved by taking the highest score on any of the four items (sleep, leisure/social, housework/errands, and work/school). For the distribution domain, the number of affected body parts is tallied (potential sum 0 – 16, where body parts with missing responses are considered not affected and assigned a value of zero for summation) and the sum is sorted into five scoring bins: sum of 0 – 2 = score of 1, sum of 3 – 5 = score of 2, sum of 6 – 10 = score of 3, sum of 11 – 13 = score of 4, and sum of 14 – 16 = score of 5. The 5D-Itch total score is then the summation of the five domain scores. 5D-Itch scores can potentially range between 5 (no pruritus) and 25 (most severe pruritus) {Elman 2010}. For missing total score, it will be imputed after the summation.

Chronic Liver Disease Questionnaire (CLDQ):

CLDQ scores are calculated using participant responses to 29 questions in the questionnaire. The 6 domain scores are calculated as follows:

- Abdominal symptoms = Mean of items {1, 5, 17}
- Fatigue = Mean of items {2, 4, 8, 11, 13}
- Systemic symptoms = Mean of items {3, 6, 21, 23, 27}
- Activity = Mean of items {7, 9, 14}
- Emotional function = Mean of items {10, 12, 15, 16, 19, 20, 24, 26}
- Worry = Mean of items {18, 22, 25, 28, 29}

Each domain score is calculated as the average of nonmissing items. Each score is calculated only if at least half of corresponding items are not missing. Otherwise, the score will be missing.

Overall CLDQ score is calculated by taking the average of the above 6 domain scores. Overall CLDQ score will be summarized.

Health Resource Utilization Questionnaire (HRUQ):

The number of visits/days for the following questions will be summarized:

- Outpatient non-emergent clinic visits (number of visits)
- Outpatient non-emergent clinic visits related to PSC (number of visits)
- Outpatient emergent department visits (number of visits)
- Outpatient emergent department visits related to PSC (number of visits)
- Outpatient surgeries and procedures (number of visits)
- Outpatient surgeries and procedures related to PSC (number of visits)
- Inpatient surgeries and procedures (number of visits)
- Inpatient surgeries and procedures related to PSC (number of visits)
- Number of days hospitalized (number of days)
- Number of days hospitalized related to PSC (number of days)

Appendix 4. Predictive Power (PP) Calculation

Predictive power calculates the probability of observing a statistically significant result of the primary analysis given the interim data, where a stratified Mantel-Haenszel test will be used for the primary analysis with adjustment for the baseline UDCA use (yes versus no) and fibrosis stage (Ludwig fibrosis score, F3 versus F0-F2) on screening liver biopsy.

Let $\tilde{n}_{jks}$ be the number of participants in the FAS enrolled up to interim ($k = 1$) and post-interim ($k = 2$) in treatment $j = 0, 1$ and stratum $s = 1,...,4$, and among them, n_{jks} be the number of participants with non-missing value at Week 96 and without liver-related clinical events prior to Week 96. Let M_{jks} be the number of responses among the n_{jks} participants.

CCI

Given observed number of responses m_{j1s} for treatment $j = 0, 1$ and stratum $s = 1,...,4$ at interim, PP is calculated by exhausting all combinations of m_{j2s} that lead to a statistically significant result; that is,

$$PP(\{m_{j1s}: j = 0, 1; s = 1,...,4\}) = \sum_{\left\{m_{j2s};\ 0 \le m_{j2s} \le n_{j2s}\right\}}^{j=0,1; s=1,...,4;} \left\{\prod_j \prod_s P\left(M_{j2s} = m_{j2s} \,\middle|\, M_{j1s} = m_{j1s}\right)\right\} \mathrm{I}(\text{pval} \le 0.025),$$

where pval is the one-sided p-value of the stratified Mantel-Haenszel test at the primary analysis as described in Section 6.1.2 to 6.1.4.

For this calculation, we assume the proportion of participants with either liver-related clinical events by Week 96 or missing value at Week 96 is the same across treatment groups and strata for all participants in the FAS, and such proportion will be estimated based on the data observed at interim.

The operating characteristics of using the proposed futility criterion of PP ≤ 10% are shown in Table 12-3 below based on Chi-square test. When the true effect size is null, the probability of meeting the futility criterion (PP ≤ 10%) to stop the trial at interim analysis is 58.9%; when the true effect size increases to 15% and 20%, the probability of meeting the futility criterion decreases to 11.2% and 4.4%, respectively. The evaluation shows that the chosen criterion

reasonably supports a correct decision with a high probability of stopping the trial at interim analysis when the true treatment effect is null, and a low probability of stopping the trial when there is a desirable treatment effect.

Table 12-3. Operating Characteristics of Futility Criterion with Predictive Power ≤ 10%

Futility Criterion (PP=Probability of success based on observed data)	True Treatment Effect	Probability of Meeting Futility Criterion
PP ≤ 10%	0%	58.9%
	15%	11.2%
	20%	4.4%

Appendix 5. Determining Missing and Virtual visits due to COVID-19

This appendix describes the clinical trial site collection of COVID-19 data pertaining to missed/virtual visits and the data processing algorithm used to determine which visits were missing and which visits were virtual.

Data collection

A COVID-19 supplement to the eCRF Completion Guidelines (CCG) was provided by data management to instruct clinical trial sites with respect to data entry expectations pertaining to scenarios related to the COVID-19 pandemic. If a visit was missed, sites should enter "Visit missed due to COVID-19." If an in-person visit was conducted virtually, sites should enter "Virtual visit due to COVID-19."

Determination of Missed and Virtual visits

Natural Language Processing (NLP) was used to search the CRF comment fields to identify instances of "COVID-19" (or synonyms, see Table 12-4) and "Virtual" (or synonyms, see Table 12-4). The search terms are maintained in a global lookup table and can be modified to tune the NLP model. For any comments with COVID-19 search terms, assign "Missed visit" or "Virtual visit as follows:

If COVID-19 terms are identified through NLP and the visit date is missing, then result is "Missed Visit"

If COVID-19 and Virtual terms are identified through NLP for a visit, then result is "Virtual Visit". When there are multiple records for the same participant and the same visit, NLP will be based on multiple records to ensure 1 unique category per participant per visit

Otherwise result is missing

Table 12-4. Examples of search terms for "COVID-19" and "Virtual" used to identify missed and virtual visits.

Search terms for "COVID-19"	Search terms for "Virtual"
COVID19	VIRTUAL
CORONA	TELEMED
CORONAVIRUS	TELEHEALTH
PANDEMIC	TELEPHONE
OUTBREAK	REMOTE
CRISIS	TELEMEDICINE
LOCKDOWN	TELECONSULTATION
QUARANTINE	TELEPHONICALLY
SHELTER	PHONE
	HOME VISIT
	ZOOM
	SKYPE

GS-US-428-4194-SAP-v1.0

ELECTRONIC SIGNATURES

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:mm:ss)
PPD	Biostatistics eSigned	12-Aug-2022 14:47:33
PPD	Clinical Research eSigned	12-Aug-2022 16:15:41