STATISTICAL ANALYSIS PLAN

Study Title:	A Phase 2, Randomized, Double-Blind, Double-Dummy, Placebo-Controlled Study Evaluating the Safety and Efficacy of Semaglutide, and the Fixed-Dose Combination of Cilofexor and Firsocostat, Alone and in Combination, in Subjects with Compensated Cirrhosis (F4) due to Nonalcoholic Steatohepatitis (NASH)
Name of Test Drug:	Semaglutide, Fixed-Dose Combination of Cilofexor and Firsocostat
Study Number:	GS-US-454-6075
Protocol Version (Date):	Amendment 2: 26 October 2023 Administrative Amendment 1: 08 January 2024
Analysis Type:	Final Analysis
Analysis Plan Version:	Version 1.0
Analysis Plan Date:	07 January 2025
Analysis Plan Author(s):	PPD

CONFIDENTIAL AND PROPRIETARY INFORMATION

TABLE OF CONTENTS

LIST OF IN-TEXT TABLES

LIST OF IN-TEXT FIGURES

LIST OF ABBREVIATIONS

AE	adverse event
ALP	alkaline phosphatase
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
APRI	AST to platelet ratio index
AST	aspartate aminotransferase
BLQ	below the limit of quantitation
BMI	body mass index
CAP	controlled attenuation parameter
CI	confidence interval
CILO	cilofexor (GS-9674)
CIR	crude incidence rate
CK	creatine kinase
COVID-19	coronavirus disease 2019
CP	Child-Pugh
CRF	case report form
CRN	Clinical Research Network
CRP	C-reactive protein
CSR	clinical study report
CTCAE	Common Terminology Criteria for Adverse Events
DMC	data monitoring committee
EAIR	exposure-adjusted incidence rate
ECG	electrocardiogram
eCRF	electronic case report form
eGFR	estimated glomerular filtration rate
ELF	enhanced liver fibrosis
ET	early termination
FACIT	Functional Assessment of Chronic Illness Therapy
FAS	Full Analysis Set
FDC	fixed-dose combination
FIB-4	Fibrosis-4
FIR	firsocostat (GS-0976)
FU	Follow-Up
FWER	Family-wise type 1 error rate
GGT	gamma-glutamyl transferase
HbA1c	hemoglobin A_{1c}
HCC	hepatocellular carcinoma

HDL	high density lipoprotein
HE	hepatic encephalopathy
HLGT	high-level group term
HLT	high-level term
HOMA-IR	homeostasis model assessment of insulin resistance
INR	international normalized ratio
IRT	Interactive Response Technology
LDL	low-density lipoprotein
LLT	lower-level term
LOCF	Last Observation Carried Forward
LOQ	limit of quantitation
MAR	missing at random
MCMC	Markov Chain Monte Carlo
MDRD	Modification of Diet in Renal Disease
MedDRA	Medical Dictionary for Regulatory Activities
MELD	Model for End-stage Liver Disease
MH	Mantel-Haenszel
MI	Multiple Imputation
MNAR	Missing not at random
MST	medical search term
NAFLD	Nonalcoholic fatty liver disease
NAS	(NAFLD) Activity Score
NASH	nonalcoholic steatohepatitis
NFS	NAFLD fibrosis score
NRI	Non-responder Imputation
OC	Observed Case
PGI-C	Patient Global Impression Scale of Change
PGI-S	Patient Global Impression Scale of Severity
PK	pharmacokinetic(s)
PRO	patient-reported outcome
PT	preferred term
PYE	patient-years of exposure
Q1, Q3	first quartile, third quartile
RBI	Reference-based Imputation
SAEs	serious AEs
SEMA	semaglutide
SAP	statistical analysis plan
SD	standard deviation
SE	standard error
SF-36	Short Form 36 Health Survey Questionnaire

SOC	system organ class
TEAE	treatment-emergent adverse event
TELA	treatment-emergent lab abnormality
TFLs	tables, figures, and listings
ULN	upper limit of normal
VLDL	very low-density lipoprotein
WHO	World Health Organization

1. INTRODUCTION

This statistical analysis plan (SAP) describes the statistical analysis methods and data presentations to be used in tables, figures, and listings (TFLs) in the clinical study report (CSR) for Study GS-US-454-6075. This SAP is based on study protocol amendment 2 dated 26 October 2023 (administrative amendment 1 dated 08 January 2024). The SAP will be finalized before database finalization. Any changes made after the finalization of the SAP will be documented in the CSR.

1.1. Study Objectives

The primary objective of this study is as follows:

To evaluate whether the combination of semaglutide (SEMA) with the fixed-dose combination (FDC) of cilofexor and firsocostat (CILO/FIR) causes fibrosis improvement (according to the NASH Clinical Research Network [CRN] classification) without worsening of NASH (defined as a ≥ 1-point increase in hepatocellular ballooning or lobular inflammation) in participants with compensated cirrhosis due to NASH, as compared with placebo.

The secondary objectives of this study are as follows:

- To confirm the contribution of CILO/FIR to fibrosis improvement without worsening of NASH in participants treated with the combination of SEMA and CILO/FIR by comparing with participants treated with SEMA alone.
- To evaluate whether the combination of SEMA with the FDC of CILO/FIR causes NASH resolution (defined as lobular inflammation of 0 or 1 and hepatocellular ballooning of 0) in participants with compensated cirrhosis due to NASH, as compared with placebo.
- To confirm the contribution of SEMA to NASH resolution in participants treated with the combination of SEMA and CILO/FIR by comparing with participants treated with CILO/FIR alone.

CCI

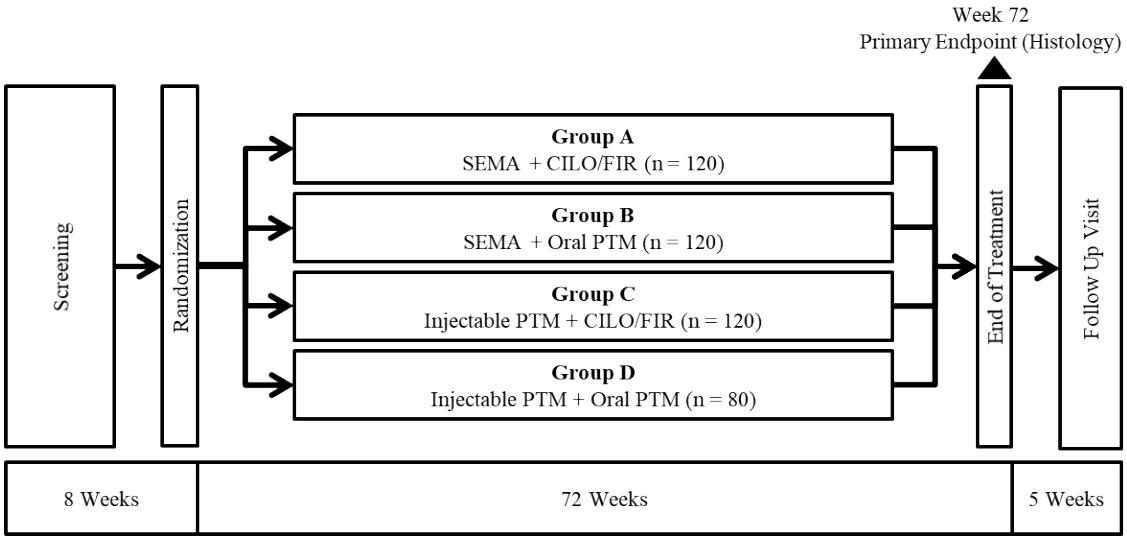
[redacted]

1.2. Study Design

This is a Phase 2, randomized, double-blind, double-dummy, placebo-controlled study evaluating the efficacy and safety of SEMA, CILO/FIR, and their combination in participants with compensated cirrhosis due to NASH.

Participants meeting the study's entry criteria will be randomly assigned in a 3:3:3:2 ratio to 1 of 3 active treatment groups (SEMA+CILO/FIR, SEMA+PTM CILO/FIR, PTM SEMA+CILO/FIR) or placebo (PTM SEMA+PTM CILO/FIR), with approximately 120 participants in each of the 3 active treatment groups and approximately 80 participants in the placebo group (PTM SEMA+PTM CILO/FIR), as shown in Figure 1.

Figure 1. GS-US-454-6075 Study Schema



Central randomization is used. Randomization will be stratified by the presence or absence of type 2 diabetes as determined by medical history or based on screening laboratory values if previously undiagnosed (ie, $HbA_{1c} \geq 6.5\%$ or fasting plasma glucose ≥ 126 mg/dL, confirmed on repeat testing), and by ELF score (≥ 11.30 or < 11.30 during screening). For the entire duration of the study, participants and investigators will remain blinded to treatment assignment.

Participants will be treated for 72 weeks. Total study duration will be up to 85 weeks, including up to 8 weeks for screening, a 72-week treatment period, and a 5-week follow-up period.

The final analysis will be conducted at the end of the study, ie, completion of the follow-up period.

1.3. Sample Size and Power

The power calculation is based on the estimated proportion of participants in the SEMA+CILO/FIR group who achieve the primary endpoint at Week 72 compared with the placebo group. Assuming the proportion of participants in the SEMA+CILO/FIR and placebo groups who achieve ≥ 1-stage fibrosis improvement without worsening of NASH at Week 72 is 35% and 12%, respectively, with a sample size of 120 participants in the SEMA+CILO/FIR group and 80 participants in the placebo group, the study has 97% power to detect a difference at a 2-sided significance level of 0.05.

If the primary endpoint is achieved, the contribution of CILO/FIR to ≥ 1-stage fibrosis improvement without worsening of NASH will be evaluated by comparing the SEMA+CILO/FIR and SEMA groups for this endpoint. Assuming the proportion of participants who achieve ≥ 1-stage fibrosis improvement without worsening of NASH at Week 72 is 35% and 20% in the SEMA+CILO/FIR and SEMA groups, respectively, with a sample size of 120 participants in each group, the study has 74% power to detect a difference at a 2-sided significance level of 0.05.

If the contribution of CILO/FIR to ≥ 1-stage fibrosis improvement without worsening of NASH is demonstrated, the overall effect of NASH resolution will be evaluated by comparing the SEMA+CILO/FIR group versus the placebo group. Assuming the proportion of participants in the SEMA+CILO/FIR and placebo groups who achieve NASH resolution at Week 72 is 45% and 10%, respectively, the study has > 99% power to detect a difference at a 2-sided significance level of 0.05.

If the superiority of SEMA+CILO/FIR versus placebo for the overall effect of NASH resolution is achieved, the contribution of SEMA to NASH resolution will be evaluated by comparing the SEMA+CILO/FIR group versus the CILO/FIR group for this endpoint. Assuming the proportion of participants who achieve NASH resolution at Week 72 is 45% and 15% in the SEMA+CILO/FIR and CILO/FIR groups, respectively, with a sample size of 120 participants in each group, the study has 99% power to detect a difference at a 2-sided significance level of 0.05.

The power calculations described above are based on Pearson's chi-square test using a normal approximation.

2. TYPE OF PLANNED ANALYSIS

2.1. Interim Analyses

No formal interim efficacy analysis is planned.

2.1.1. DMC Interim Safety Review

An external multidisciplinary Data Monitoring Committee (DMC) will review the progress of the study and perform interim, and as necessary, ad hoc 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 review will be conducted after at least 55 participants have completed through Week 4. Additional meetings will be scheduled approximately every 6 months thereafter.

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.

In the course of ongoing safety monitoring, if at any point any of the below criteria are met, an urgent meeting of the DMC will be held to determine whether the study should proceed as planned, proceed with modification, or be terminated:

- ≥ 2 participants develop CTCAE Grade 4 pancreatitis (life threatening) deemed related to study drug(s) by the investigator
- ≥ 2 participants develop the same CTCAE Grade 4 unexpected AE (by preferred term) deemed related to study drug(s) by the investigator
- Any participant develops a CTCAE Grade 5 unexpected AE (death) deemed related to study drug(s) by the investigator

While the DMC will be asked to advise Gilead regarding future conduct of the study, including possible early study termination, Gilead retains final decision-making authority on all aspects of the study.

2.2. Final Analysis

The unblinded final analysis will be performed after all participants have completed 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 2-sided and performed at the 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 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 randomized 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, if applicable.

For each analysis set, the number and percentage of participants eligible for inclusion will be summarized by treatment group.

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

3.1.1. All Randomized Analysis Set

The 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 received 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 received at least 1 dose of study drug. This is the primary analysis set for safety analyses.

3.1.4. Pharmacokinetic Analysis Set

The Pharmacokinetic (PK) Analysis Set includes all randomized participants who received at least 1 dose of study drug and have at least 1 nonmissing postbaseline concentration value reported by the PK laboratory for analytes CILO, FIR, and their respective metabolites, as applicable. This is the primary analysis set for all PK analyses and will be used for the listing of concentration data.

3.2. Participant Grouping

For analyses based on the All Randomized Analysis Set and FAS, participants will be grouped according to the treatment to which they were randomized.

For analyses based on the Safety 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 this case, the actual treatment received is defined as the treatment received for the entire treatment duration.

For the PK Analysis Set, participants will be grouped according to the actual treatment they received.

3.3. Strata and Covariates

Participants will be randomly assigned to treatment groups via the interactive response technology (IRT) in a 3:3:3:2 ratio to 1 of 3 active treatment groups (SEMA+CILO/FIR, SEMA+PTM CILO/FIR, PTM SEMA+CILO/FIR) or placebo (PTM SEMA+PTM CILO/FIR) using a stratified randomization schedule. Stratification will be based on the following variables:

- Presence or absence of type 2 diabetes as determined by medical history or based on screening laboratory values if previously undiagnosed (ie, $HbA_{1c} \geq 6.5\%$ or fasting plasma glucose ≥ 126 mg/dL, confirmed on repeat testing)
- ELF test score ≥ 11.30 or < 11.30 during screening

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 covariates or stratification variables for analyses, as specified in Section 6.

For efficacy endpoints, the baseline value of the efficacy variable(s) will be included as a covariate in the efficacy analysis model, if applicable.

3.4. Examination of Participant Subgroups

The primary and secondary efficacy endpoints, unless otherwise specified, will be examined using the following subgroups in primary analysis:

- Participants with diabetes at baseline
- Participants without diabetes at baseline
- Participants with ELF score ≥ 11.30 at baseline

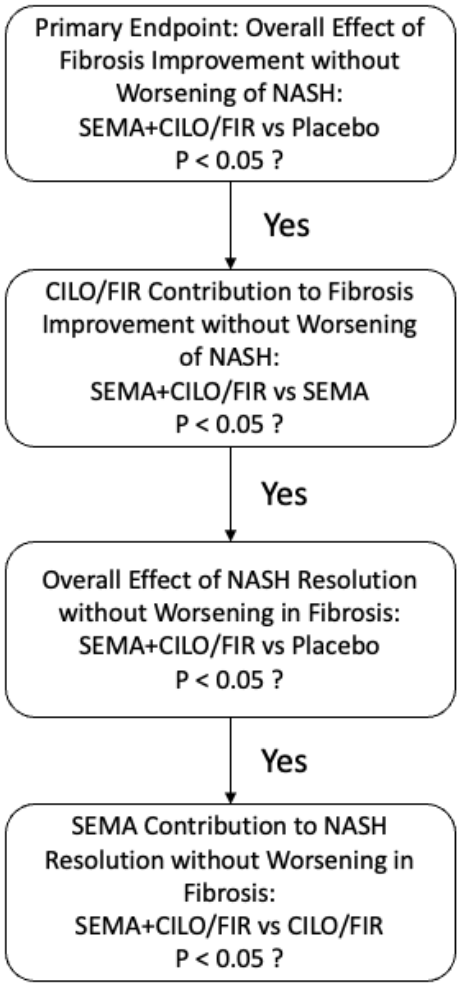
- Participants with ELF score < 11.30 at baseline
- Participants with baseline CRN fibrosis stage = 4 or ≤ 3 separately based on re-read of baseline slides at the time of final analysis for the following endpoints:
  - ≥ 1-stage improvement in fibrosis without worsening of NASH at Week 72 in the SEMA+CILO/FIR versus placebo groups (Section 6.1.1)
  - ≥ 1-stage improvement in fibrosis without worsening of NASH at Week 72 in the SEMA+CILO/FIR versus SEMA groups (Section 6.2.1)
- Participants with baseline CRN fibrosis stage = 4, hepatocellular ballooning grade ≥ 1, and lobular inflammation grade ≥ 1 based on re-read of baseline slides at the time of final analysis

For the comparison of treatment difference in each subgroup, the point estimate and 95% confidence interval (CI) based on the analysis methods for the primary and secondary endpoints as described in Sections 6.1 and 6.2, respectively, will be presented. For each endpoint, if the evaluable sample size after missing data imputation is less than 2 in a treatment group in a stratum based on baseline diabetes status and ELF score (≥ 11.30 or < 11.30), the stratification factor of baseline diabetes status will be excluded from the consideration. If the evaluable sample size is still less than 2 in a treatment group in a stratum based on baseline ELF score, an unstratified version of the Mantel-Haenszel (MH) methods described in Section 6.1.4 will be applied.

3.5. Multiple Comparisons

The multiple testing strategy is presented graphically in Figure 2 and is described below for the primary histologic endpoint and the secondary histologic endpoints.

Figure 2. Multiple Testing Strategy



The family-wise type I error rate (FWER) will be controlled through the following sequential testing procedure at a 2-sided significance level of 0.05 (equivalent to one-sided 0.025).

First, the primary endpoint will be tested at a 2-sided significance level of 0.05. The primary objective will be considered achieved if the superiority of SEMA+CILO/FIR versus placebo for the primary endpoint as described below is met.

The overall effect of the primary histologic efficacy endpoint of fibrosis improvement without worsening of NASH will be compared between the SEMA+CILO/FIR and placebo groups at Week 72 at a 2-sided significance level of 0.05. Superiority of SEMA+CILO/FIR versus placebo will be demonstrated through the test of the null hypothesis of equal responder proportions against the alternative hypothesis of a higher proportion in the SEMA+CILO/FIR group.

If the primary objective is achieved, the secondary efficacy endpoints will be tested sequentially in the following order at a 2-sided significance level of 0.05. If superiority is achieved for a secondary endpoint, the next secondary endpoint will be evaluated; otherwise, testing of the remaining secondary endpoints will cease.

1) The contribution of CILO/FIR to ≥ 1-stage fibrosis improvement without worsening of NASH will be evaluated by comparing the SEMA+CILO/FIR and SEMA groups for this endpoint at a 2-sided significance level of 0.05. Superiority of SEMA+CILO/FIR versus SEMA will be demonstrated through the test of the null hypothesis of equal responder proportions against the alternative hypothesis of a higher proportion in the SEMA+CILO/FIR group.

2) The overall effect of NASH resolution without worsening in fibrosis will be compared between the SEMA+CILO/FIR and placebo groups at Week 72 at a 2-sided significance level of 0.05. Superiority of SEMA+CILO/FIR versus placebo will be demonstrated through the test of the null hypothesis of equal responder proportions against the alternative hypothesis of a higher proportion in the SEMA+CILO/FIR group.

3) The contribution of SEMA to NASH resolution without worsening in fibrosis will be evaluated by comparing the SEMA+CILO/FIR and CILO/FIR groups for this endpoint at a 2-sided significance level of 0.05. Superiority of SEMA+CILO/FIR versus CILO/FIR will be demonstrated through the test of the null hypothesis of equal responder proportions against the alternative hypothesis of a higher proportion in the SEMA+CILO/FIR group.

3.6. Missing Data and Outliers

3.6.1. Missing Data

In general, missing data will not be imputed unless methods for handling missing data are specified. Exceptions are presented in this document.

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 7.1.7.2, and for prior and concomitant medications in Section 7.4.

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

1) Nonresponder Imputation (NRI)

   For the purpose of analysis at a defined time point, when NRI is used, participants with missing values will be analyzed as treatment failure.

2) Reference-based Imputation (RBI) assuming Missing Not at Random (MNAR):

   When the endpoint is binary, missing data in all treatment groups will be imputed from logistic regression with stratification factors, ie, baseline diabetes status and ELF score (≥ 11.30 or < 11.30), as covariates using nonmissing data in the placebo group.

When the endpoint is continuous, missing data in all treatment groups will be imputed from the imputation model using nonmissing data in the placebo arm, with stratification factors (and other predictor variables as appropriate) as covariates. Intermittent missing values will be imputed by visit using Markov Chain Monte Carlo (MCMC) which assumes a multivariate normal distribution over all variables included in the imputation model. Intermittent missing data will first be multiply imputed assuming missing at random and fitting a monotone missing pattern. Then the remaining monotone missing values will be multiply imputed using the placebo group profile (ie, reference-based imputation) using PROC MI with MNAR statement.

A total of 100 imputed datasets will be generated by the imputation method. That is, each missing data point will be imputed 100 times. Each imputed data set will be analyzed by the method for the primary, secondary, and CCI [redacted]. The results from these 100 imputed data sets will be combined using Rubin's rule {Rubin 1987}.

3) Multiple Imputation (MI) assuming Missing at Random (MAR):

For the primary endpoint, 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 missing at random (MAR) assumption. The imputation model used to impute missing data points will be based on the logistic regression with the treatment arms and the randomization stratification factors as covariates. A total of 100 imputed datasets will be generated by the imputation model. That is, each missing data point will be imputed 100 times. Each imputed data set will be analyzed by the method for the primary analysis as specified in Section 6.1.4. The results from these 100 imputed data sets will be combined using Rubin's rule {Rubin 1987}.

4) Tipping Point Analysis (TPA):

If the primary analysis of the primary endpoint is significant in favor of SEMA+CILO/FIR, a delta-adjusting pattern-mixture approach for TPA {Ratitch 2013} will be conducted to assess the robustness of the primary analysis results under a missing not at random assumption (MNAR) for the SEMA+CILO/FIR and placebo arms. Specifically, it is assumed that a systematic difference exists between the conditional distributions of the missing and observed data in the SEMA+CILO/FIR or placebo arm. To reflect such a systematic difference, two shift parameters δ_1 and δ_2 will be applied to the imputation model (ie, the logistic regression with the treatment arms and the randomization stratification factors as covariates) when the missing data points are imputed. The values of parameter δ_1 will be unfavorable for the SEMA+CILO/FIR arm and that of δ_2 will be favorable for the placebo arm, ie, reducing the probability of being a responder for the SEMA+CILO/FIR participants and increasing the probability of being a responder for the placebo participants. For each value of δ_1 and δ_2, multiply 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_1 and δ_2, the impact from missing data on the analysis results will be examined to identify tipping points (ie, the values at which conclusions from statistical inference change from being significant to being insignificant for the SEMA+CILO/FIR arm).

5) Observed Case (OC)

   Only participants with observed baseline (baseline re-read for histological endpoints based on pathologists' reading) and Week 72 values will be included at the final analysis.

6) Last Observation Carried Forward (LOCF)

   For the 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. The LOCF approach will be applied to those participants who have a baseline value and at least one postbaseline value at an analysis visit.

The RBI (or NRI when appropriate) approach under estimand framework in Section 6.1.2 and 6.2.2 will be used for primary analyses for the primary and secondary efficacy endpoints.

Nonresponder imputation (if not applied per the primary analysis estimand framework) 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.

The MI and TPA approaches will be carried out as sensitivity analyses of the primary efficacy endpoint if applicable.

CCI [redacted]

3.6.2. Outliers

Outliers will be identified during the data management and data analysis process, but no sensitivity analyses will be conducted. All data will be included in the data analysis.

3.7. Data Handling Conventions and Transformations

Only year of birth is collected on the CRF; "01July" will be imputed as the day and month of birth.

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 was signed will be used for the age derivation. Age required for longitudinal and temporal calculations and analyses (eg, 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 lower LOQ will be used to calculate descriptive statistics if the datum is reported in the form of "< x" (where x is considered the lower LOQ). For example, if the values are reported as < 50 or < 5.0, values of 49 or 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 upper LOQ will be used to calculate descriptive statistics if the datum is reported in the form of "> x" (where x is considered the upper 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 III N-terminal propeptide [PIIINP], and tissue inhibitor of metalloproteinase 1 [TIMP1]) is less than the lower LOQ or above the upper LOQ, the ELF components 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 6.

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

3.8. Analysis Visit Windows

3.8.1. Definition of Study Day

Study day will be calculated from the first dosing date of study drug(s) and derived as follows:

- For postdose study days: Assessment Date – First Dosing Date + 1

- For days prior to the first dose: Assessment Date – First Dosing Date

Therefore, Study Day 1 is the day of first dose of study drug administration.

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.

On-treatment visit windows will be calculated from Study Day 1 for selected efficacy measures, vital signs, and safety laboratory data. Except for histologic biopsy data or liver stiffness and CAP by transient elastography data, selected safety and efficacy data, collected up to and including the last dosing date + 35 days for participants who have permanently discontinued study drug, or the database snapshot date for participants who were still on treatment at the time of analysis for DMC safety review, will be mapped according to the following analysis windows unless the nominal visit name is Follow-Up (FU).

The analysis windows for selected measures are provided in Table 1, Table 2, Table 3, Table 4 and Table 5. The algorithm for assigning baseline analysis window does not apply to the liver tests (ALP, total bilirubin, direct bilirubin, ALT, AST, and GGT) and LDL. For these parameters, the baseline values will be determined by averaging the values obtained between the Screening and Day 1 (inclusive), and no specific study day is associated with the average values.

Table 1. Analysis Visit Windows for Chemistry, Hematology, Coagulation, CP/MELD scores, Lipids, eGFR, Vital Signs, and Body Weight

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	98
Week 16	113	99	140
Week 24	169	141	210
Week 36	253	211	294
Week 48	337	295	378
Week 60	421	379	462
Week 72	505	463	≥505

Table 2. Analysis Visit Windows for Glycemic Panel, HbA1c, ELF™ Score, Hyaluronic Acid, PIINP, TIMP1, FibroSure/FibroTest and other Noninvasive Tests (NITs)

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	≥505

Table 3. Analysis Visit Windows for Hip and Waist Circumference, Liver Stiffness and CAP by Transient Elastography, and ECG

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	≥505

Table 4. Analysis Visit Windows for Liver Biopsy

Nominal Visit	Nominal Study Day	Visit Window Study Day	
		Lower Limit	Upper Limit
Baseline	1	(none)	1
Week 72	505	337	≥505

Table 5. Analysis Visit Windows for PROs

Nominal Visit	Nominal Study Day	Visit Window Study Day	
		Lower Limit	Upper Limit
Baseline	1	(none)	1
Week 24	169	155	183
Week 48	337	323	351
Week 72	505	491	519

Any data relating to unscheduled visits and early termination visit will be assigned to a particular visit or time point based on the analysis visit window. However, the following conventions will be used:

- 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.
- For participants who prematurely discontinue from the study, early termination (ET) data will be assigned to a visit window depending on the study day of the visit for parameters to be windowed, except for PROs. ET data will be summarized as a separate visit for PROs, labeled as "Early Termination Visit".
- Data collected for a follow-up visit will be summarized as a separate visit and labeled "Follow-up Visit."
- Data obtained after the follow-up visit or last dose date plus 35 days for participants who have permanently discontinued study drug (whichever is later) will be excluded from the summaries except for histologic biopsy data or liver stiffness and CAP by transient elastography data but will be included in the listings.

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

For the stratification factor of presence or absence of type 2 diabetes based on screening laboratory values when previously undiagnosed, ie, $HbA_{1c} \geq 6.5\%$ or fasting plasma glucose ≥ 126 mg/dL confirmed on repeat testing, if any of the following 2 conditions is met, the stratification factor of type 2 diabetes is "presence"; otherwise, "absence":

- At least 2 records (consecutive or not) between screening and Day 1 (inclusive) have HbA1c ≥ 6.5%
- At least 2 records (consecutive or not) between screening and Day 1 (inclusive) have fasting plasma glucose ≥ 126 mg/dL

For the stratification factor of ELF test score ≥ 11.30 or < 11.30 during screening, the stratification factor of ELF test score is based on the latest ELF score between screening and Day 1 (inclusive).

Depending on the statistical analysis method, single values 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:

- For baseline, the last nonmissing value on or prior to the first dosing date of study drug will be selected, unless specified differently. If there are multiple records on the same day, the baseline value will be the average of the measurements for continuous data, or the measurement with the lowest severity (eg, normal will be selected over abnormal for safety electrocardiogram [ECG] findings) for categorical data.
- Baseline values for ALP, ALT, AST, total bilirubin, direct bilirubin, GGT, and LDL will be calculated by averaging the values obtained between and including screening and Day 1. If the participant was rescreened, data from previously failed screening will be excluded from the calculation. 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 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 worse severity will be taken for categorical data, unless otherwise specified.

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

- For baseline, measurements by XL probe will be selected for analysis if available, otherwise measurements by M probe will be selected.
- For postbaseline visits, measurements by the same probe type (XL or M) 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 analysis value for the corresponding postbaseline visit will be considered missing. If there are multiple postbaseline records by the same probe type as baseline, the rules to choose postbaseline continuous measurements as described above will apply.

Histologic biopsy data based on pathologists' reading in each analysis visit window will be chosen based on the following rules:

- For original baseline reading, the latest nonmissing reading at screening will be selected for analysis
- For baseline re-read, the latest nonmissing reading of the baseline biopsy will be selected for analysis
- For Week 72, the latest nonmissing reading of Week 72 biopsy will be selected for analysis

Histologic biopsy data based on PathAI reading in each analysis visit window will be chosen based on the following rules:

- For baseline, the latest nonmissing record of baseline biopsy will be selected for analysis
- For Week 72, the latest nonmissing record of Week 72 biopsy will be selected for analysis.

4. PARTICIPANT DISPOSITION

4.1. Participant Enrollment and Disposition

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

A summary of participant enrollment will be provided by treatment group for each country, investigator within a country, and overall. 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 overall. This summary will present the number of participants screened, the number of participants randomized, and the number of participants in each of the categories listed below (as applicable):

- Full Analysis Set
- Pharmacokinetic Analysis Set
- Safety Analysis Set
- Completed both study drugs
- Did not complete study drug with reasons for premature discontinuation of any study drug
- Completed study
- Did not complete the study with reasons for premature discontinuation of study

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 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

For the reasons for premature discontinuation of any study drug, if either reason for CILO/FIR (or PTM) or SEMA (or PTM) discontinuation is AE, then AE will be used in the summary of reason for premature discontinuation of any study drug.

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

The last dose date of individual study drug will be calculated separately for each study drug in a treatment group and will be the end date on the study drug administration case report form (CRF) for the record where the "study drug was permanently withdrawn" flag is 'Yes'.

Total duration of exposure to individual study drug that is administered daily (ie, CILO/FIR) will be defined as last dose date of individual study drug minus first dose date of individual study drug plus 1. Total duration of exposure to individual study drug that is administered weekly (ie, SEMA) will be defined as last dose date of individual study drug minus first dose date of individual study drug plus 7. Total duration of exposure to treatment regimen will be defined as the maximum of [(last dose date of SEMA plus 6) and last dose date of CILO/FIR] minus the first dose date of any study drug plus 1. Total duration of exposure will be calculated 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).

- Total duration of exposure to treatment regimen (weeks) = [maximum of (last dose date of SEMA + 6, last dose date of CILO/FIR) – first dose date of any study drug + 1] / 7
- Total duration of exposure to daily administered individual study drug (weeks) = (last dose date of individual study drug – first dose date of individual study drug + 1) / 7
- Total duration of exposure to weekly administered individual study drug (weeks) = (last dose date of individual study drug – first dose date of individual study drug + 7) / 7

If the last study drug dose 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 will be used for participants included in the final analysis. At the time of any interim analysis (ie, for DMC), the missing last dosing date will be imputed by the data snapshot date for participants who are still on treatment. 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 treatment regimen and the total duration of exposure to each individual study drug will be summarized using descriptive statistics. In addition, the total duration of exposure to treatment regimen will be summarized by the number (ie, cumulative counts) and percentage of participants exposed through the following time periods: 1 day, 4 weeks, 8 weeks, 12 weeks, 16 weeks, 20 weeks, 24 weeks, 36 weeks, 48 weeks, 60 weeks, and 72 weeks. Summaries will be provided by treatment group for the Safety Analysis Set.

No formal statistical testing is planned.

4.2.2. Adherence to Study Drug

For CILO/FIR, which is orally administered once daily in tablet form, the presumed total amount of study drug administered to a participant will be determined by the data collected on the drug accountability CRF using the following formula:

$$\text{Total Amount of Study Drug Administered} = \left(\sum \text{No. of Tablets Dispensed}\right) - \left(\sum \text{No. of Tablets Returned}\right)$$

For SEMA, which is injected once weekly, the presumed total amount of study drug administered to a participant will be determined by the data collected on the study drug administration CRF using the following formula:

$$\text{Total Amount of Study Drug Administered (mg)} = \sum \text{Actual Dose (mg)}$$

In addition, the presumed total number of study drug injection weeks will be determined by the data collected on the study drug administration CRF.

The total amount of study drug administered and total number of study drug injection weeks (for SEMA only) will be summarized using descriptive statistics for each individual study drug by treatment group for the Safety Analysis Set.

4.2.2.1. On-treatment Adherence

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:

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

In addition, the level of on-treatment adherence based on the number of injection weeks for SEMA will be expressed as a percentage using the following formula:

$$\text{On-Treatment Adherence (\%) by Number of Injection Weeks} = \left(\frac{\text{Total Number of Study Drug Injection Weeks}}{\text{Number of Injection Weeks Expected while on Treatment}}\right) \text{x } 100$$

Descriptive statistics for the level of on-treatment adherence with the number and percentage of participants belonging to adherence categories < 75%, ≥ 75 to < 90%, and ≥ 90% will be provided by treatment group for the Safety Analysis Set.

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 permitted by the sponsor to continue in the study 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 the 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 at least 1 important protocol deviation; with 1, 2, 3, or more important protocol deviations; and the total number of important protocol deviations by deviation category (eg, eligibility criteria, informed consent) will be

summarized by treatment group for the All Randomized Analysis Set. A by-participant listing will be provided for those participants with important protocol deviations.

4.4. Assessment of COVID-19 Impact

This section describes how special situations due to COVID-19 will be handled in the analysis, if applicable. Data collection and determination of COVID-19 impact data are in Appendix 7.

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

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

4.4.2. Protocol Deviations Related to COVID-19

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

4.4.3. Missed and Virtual Visits due to COVID-19

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.

4.4.4. Adverse Events Due to COVID-19

Adverse events of COVID-19 will be included in analyses of AEs 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.

5. BASELINE CHARACTERISTICS

5.1. Demographics and Baseline Characteristics

Participant demographic variables (ie, age, age category (< 60 and ≥ 60 years), sex at birth, race, and ethnicity) and baseline characteristics (body weight [in kg], height [in cm], body mass index [BMI; in kg/m^2], BMI category [18 to < 25 kg/m^2, 25 to < 30 kg/m^2, 30 to < 35 kg/m^2, and ≥ 35 kg/m^2]) will be summarized by treatment 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 Safety Analysis Set.

A by-participant demographic listing, including the informed consent date, will be provided by participant ID number in ascending order.

5.2. Other Baseline Characteristics

Other baseline characteristics to be summarized include the following; histologic biopsy parameters will include original baseline reading and baseline re-read from pathologist, and baseline reading from PathAI:

- NASH CRN fibrosis stage,
- Modified Ishak fibrosis stage,
- NAFLD activity score (NAS),
- Steatosis grade,
- Lobular inflammation grade,
- Hepatocellular ballooning grade,
- ALT (U/L),
- AST (U/L),
- ALP (U/L),
- GGT (U/L),
- Total bilirubin (mg/dL),
- Direct bilirubin (mg/dL),
- INR,

- Platelets ($\times 10^3/\mu L$),
- Hemoglobin (g/dL)
- CP score,
- MELD score,
- AST to platelet ratio index (APRI),
- Creatinine (mg/dL)
- eGFR by MDRD and eGFR category (30 to < 44 mL/min/1.73 m^2, 44 to < 60 mL/min/1.73 m^2, 60 to < 90 mL/min/1.73 m^2, ≥ 90 mL/min/1.73 m^2),
- Fasting total bile acids (μmol/L),
- Liver stiffness by transient elastography,
- Liver stiffness by transient elastography category (<14.0 kPa, 14.0 to < 30.7 kPa, ≥ 30.7 kPa),
- CAP from transient elastography if available,
- ELF™ score and its components (hyaluronic acid, PIIINP, and TIMP1) and ELF score category (< 11.30, ≥ 11.30),
- FibroSure/FibroTest score,
- Fibrosis-4 (FIB-4) and FIB-4 category (< 1.30, 1.30 to < 2.67, ≥ 2.67),
- NAFLD fibrosis score (NFS),
- NFS Category (< −1.455, −1.455 to < 0.676, ≥ 0.676),
- Total cholesterol (mg/dL),
- HDL cholesterol (mg/dL),
- Fasting LDL cholesterol (mg/dL),
- Fasting non-HDL cholesterol (mg/dL),
- Fasting VLDL cholesterol (mg/dL),
- Fasting triglycerides (mg/dL),

- Fasting insulin (µIU/mL),
- Fasting plasma glucose (mg/dL),
- HOMA-IR,
- HbA1c (%),
- Use of statin or other lipid-altering therapy (eg, fibrates) at baseline (yes or no),
- Use of Vitamin E at baseline (yes or no),
- Diabetes mellitus (absence or presence),
- Waist circumference (cm),
- Hip circumference (cm).

These baseline characteristics will be summarized by treatment group 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.

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

The algorithms for derivation of ELF, NFS, APRI, FIB4, and FibroTest/FibroSure are provided in Appendix 6.

5.3. Medical History

Medical history will be collected at screening for disease-specific and general conditions (ie, conditions not specific to the disease being studied). Medical history will be coded using the current version of Medical Dictionary for Regulatory Activities (MedDRA).

General and disease-specific medical history will be combined and summarized by treatment group and overall by the number and percentage of participants with each prepopulated condition for the Safety Analysis Set. No formal statistical testing is planned.

6. EFFICACY ANALYSES

Efficacy analyses will be performed after all participants have completed the study, outstanding data queries have been resolved or adjudicated as unresolvable, and the data have been cleaned and finalized.

Efficacy data will be summarized and analyzed based on the FAS. All histologic endpoints based on pathologists' reading will be based on the re-read of baseline biopsy and the reading of postbaseline biopsy by the central pathologists. Final fibrosis score and NAS score (defined in Appendix 5) will be the statistical mode of three independent pathologist reads or based on a consensus read of all three pathologists when no mode exists. For more details, refer to the Liver Biopsy Central Pathology Charter.

For continuous efficacy endpoints

Descriptive statistics will be provided by treatment group as follows:

- Baseline value
- Value at postbaseline visits
- Change from baseline at postbaseline visits
- Percent change from baseline (and percent ratio to baseline for LDL and triglycerides) at postbaseline except for MELD score, NFS, INR, HOMA-IR, CP score and PROs

CCI [redacted]

For categorical efficacy endpoints

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

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

6.1. Primary Histologic Efficacy Endpoint

6.1.1. Definition of Primary Histologic Endpoint

The primary endpoint at Week 72 is:

- $\geq$ 1-stage improvement in fibrosis (according to the NASH CRN classification) without worsening of NASH (defined as a $\geq$ 1-point increase in hepatocellular ballooning or lobular inflammation) at Week 72 in the SEMA+CILO/FIR versus placebo groups

6.1.2. Estimand for the Primary Histologic Efficacy Endpoint

The following hybrid estimand of composite and treatment policy strategies (Table 6) will be used as the estimand for the primary histologic 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 72 liver biopsy, which will be handled by a composite strategy.

Table 6. Estimand Strategy for Primary Histologic Endpoint

Primary Endpoint	Hybrid Composite/Treatment Policy Strategy
Population	All participants in the FAS as defined in Section 3.1.2
Patient Level Outcomes to be Measured	A binary response of ≥ 1-stage fibrosis improvement without worsening of NASH (defined as a ≥ 1-point increase in hepatocellular ballooning or lobular inflammation) at Week 72 in the SEMA+CILO/FIR vs placebo groups based on the baseline re-read and postbaseline biopsy reading by the central pathologists
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 72 liver biopsy are assumed to be nonresponders 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 72 liver biopsy (ie, dose interruption/reduction, protocol deviation, or premature study drug discontinuation).
Population level summary measure	Difference in proportions of participants with ≥ 1-stage fibrosis improvement without worsening of NASH between the SEMA+CILO/FIR and placebo groups.
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 72 between SEMA+CILO/FIR and placebo, adjusting for baseline diabetes status and ELF score (≥ 11.30 or < 11.30). Participants who experienced a liver-related clinical event or died due to any cause (all-cause mortality) prior to the Week 72 liver biopsy are considered nonresponders. Participants who had missing data at Week 72 due to reasons other than death or a liver-related clinical event[a] will be handled by reference-based imputation, ie, a method of multiple imputation that is informed by data from placebo participants. It is thereby assumed that participants in either treatment group without an observed outcome have the same chances of meeting the endpoint as participants in the placebo group with an observed outcome. If the number of participants with nonmissing data on this endpoint, after applying the composite and treatment policy estimand, is less than 10 in the placebo arm in at least one stratum, NRI will be used instead of RBI.

a Liver-related clinical events include evidence of progression to decompensated cirrhosis, liver transplantation, or qualification for liver transplantation on basis of MELD score (Protocol Section 7.1.4.3).

6.1.3. Statistical Hypothesis for Primary Histologic Efficacy Endpoint

The statistical hypothesis to be tested can be stated as:

$$H_0{:}\delta = 0 \text{ versus } H_1{:}\delta \neq 0$$

where δ is the difference in the proportion of participants who achieve a ≥ 1-stage improvement in fibrosis without worsening of NASH at Week 72 between SEMA+CILO/FIR and placebo.

The hypothesis for the primary endpoint will be tested in the FAS at a 2-sided significance level of 0.05. The primary objective will be considered achieved if the superiority of SEMA+CILO/FIR versus placebo for the primary endpoint is met.

6.1.4. Analysis of the Primary Histologic Efficacy Endpoint

Stratified Mantel-Haenszel (MH) tests will be used to compare the difference in proportions of participants who achieve a ≥ 1-stage improvement in fibrosis without worsening of NASH at Week 72 between SEMA+CILO/FIR and placebo, adjusting for the presence or absence of diabetes mellitus at baseline and ELF score ≥ 11.30 or < 11.30 at baseline, among eligible participants in FAS. Participants who experienced a liver-related clinical event or died due to any cause (all-cause mortality) prior to the Week 72 liver biopsy are considered nonresponders. Participant who experienced intercurrent events other than death or a liver-related clinical event prior to the Week 72 liver biopsy (ie, dose interruption/reduction, protocol deviation, or premature study drug discontinuation) will be handled by the treatment policy as described in Section 6.1.2. The missing data at Week 72 due to reasons other than death or a liver-related clinical event will be handled by RBI method. If the number of participants with nonmissing data on this endpoint, after applying the composite and treatment policy estimand, is less than 10 in the placebo arm in at least one stratum, NRI will be used instead of RBI for primary analysis.

The point estimate and 2-sided 95% CI for the difference in proportions between SEMA+CILO/FIR and placebo at Week 72 will be constructed based on stratum-adjusted MH proportions as follows {Koch 1989}:

$$p_A - p_B \pm Z_{\left(1-\frac{\alpha}{2}\right)} * SE(p_A - p_B),$$

where

- $p_A - p_B = \sum w_h d_h / \sum w_h$, is the stratum adjusted- MH proportion difference, where d_h $p_{Ah} - p_{Bh}$ is the difference in the proportion of participants who achieve a ≥ 1-stage improvement in fibrosis without worsening of NASH between SEMA+CILO/FIR and placebo in stratum h .

- $w_h = n_{Ah} n_{Bh} / (n_{Ah} + n_{Bh})$, is the weight based on the harmonic mean of sample size per treatment group for each stratum where n_{Ah} and n_{Bh} are the sample sizes of SEMA+CILO/FIR and placebo in stratum h.

- $SE(p_A - p_B) = \sqrt{\sum w_h^2 \left[\frac{p^*_{Ah}\left(1-p^*_{Ah}\right)}{n_{Ah}-1} + \frac{p^*_{Bh}\left(1-p^*_{Bh}\right)}{n_{Bh}-1}\right] \Big/ (\sum w_h)^2}$,
  where $p^*_{Ah} = (m_{Ah} + 0.5)/(n_{Ah} + 1)$ and $p^*_{Bh} = (m_{Bh} + 0.5)/(n_{Bh} + 1)$, and m_{Ah} and m_{Bh} are the number of participants who achieved a ≥ 1-stage improvement in fibrosis without worsening of NASH in SEMA+CILO/FIR and placebo in stratum h.

- α = 0.05 for the calculation of 95% CI

- $Z_{(1-\alpha/2)}$ is the 97.5th percentile of the standard normal distribution

- Z score = $\frac{(p_A - p_B)}{SE_{(p_A - p_B)}}$

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 estimate, 95% CI and corresponding *P* value based on Z tests will be provided for the difference in proportions. The 95% CIs for proportion of responders in each treatment group will be calculated through the normal approximation method for RBI and MI approaches, and Clopper-Pearson for NRI and OC approaches.

The primary efficacy objective will be achieved if $p_A > p_B$ and the 2-sided *P* value of the Z test is < 0.05 for the primary endpoint; or equivalently, if the lower confidence bound of the 95% CI is greater than 0.

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

6.1.5. Sensitivity Analysis of Primary Histologic Efficacy Endpoint

The analysis described in Section 6.1.4 will be performed after applying NRI (if not applied per the primary analysis estimand framework), MI, and OC under the hybrid of composite and treatment policy of handling intercurrent events for sensitivity analysis (see details of the imputation methods in Section 3.6.1).

If the evaluable sample size after missing data imputation is less than 2 in a treatment group in a stratum, defined by baseline diabetes status and ELF score (≥ 11.30 or < 11.30), the stratification factor of baseline diabetes status will be excluded from the consideration. If the evaluable sample size is still less than 2 in a treatment group in a stratum, as defined by baseline ELF score, an unstratified version of the MH methods described in Section 6.1.4 will be applied.

If the primary analysis of the primary endpoint is significant in favor of SEMA+CILO/FIR, the TPA approaches will be performed for sensitivity analysis.

6.2. Secondary Histologic Efficacy Endpoints

6.2.1. Definition of Secondary Histologic Efficacy Endpoints

The secondary endpoints of this study are as follows:

- ≥ 1-stage improvement in fibrosis (according to the NASH CRN classification) without worsening of NASH at Week 72 in participants treated with SEMA+CILO/FIR versus SEMA alone
- NASH resolution (defined as lobular inflammation of 0 or 1 and hepatocellular ballooning of 0) without worsening in fibrosis (if applicable) at Week 72 in the SEMA+CILO/FIR versus placebo groups
- NASH resolution without worsening in fibrosis (if applicable) at Week 72 in participants treated with SEMA+CILO/FIR versus CILO/FIR alone

6.2.2. Estimands for the Secondary Histologic Endpoints

The same estimand strategy as described for the primary endpoint will be used for the secondary endpoints. To demonstrate the contribution of CILO/FIR to fibrosis improvement without worsening of NASH, Table 7 describes the estimand strategy for the secondary endpoint of ≥ 1-stage fibrosis improvement without worsening in NASH at Week 72 in the SEMA+CILO/FIR versus SEMA groups.

Table 7. Estimand Strategy for Secondary Endpoint 1: Fibrosis Improvement Without Worsening of NASH at Week 72 in the SEMA+CILO/FIR versus SEMA Groups

Secondary Endpoint 1	Hybrid Composite/Treatment Policy Strategy
Population	All participants in the FAS as defined in Section 3.1.2
Patient Level Outcomes to be Measured	A binary response of with ≥ 1-stage fibrosis improvement without worsening of NASH at Week 72 in the SEMA+CILO/FIR versus SEMA groups, based on the baseline re-read and postbaseline biopsy reading by the central pathologists
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 72 liver biopsy are assumed to be nonresponders for the secondary 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 72 liver biopsy (ie, dose interruption/reduction, protocol deviation, or premature study drug discontinuation).
Population level summary measure	Difference in proportions of participants with ≥ 1-stage fibrosis improvement without worsening of NASH between the SEMA+CILO/FIR and SEMA groups.
Main Estimators	A stratified Mantel-Haenszel test will be used to compare the difference in proportions of participants who meet fibrosis improvement without worsening of NASH at Week 72 between the SEMA+CILO/FIR and SEMA groups, adjusting baseline diabetes status and ELF score (≥ 11.30 or < 11.30). Participants who experienced a liver-related clinical event or died due to any cause (all-cause mortality) prior to the Week 72 liver biopsy are considered nonresponders. Participants who had missing data at Week 72 due to reasons other than death or a liver-related clinical event[a] will be handled by reference-based imputation. If the number of participants with nonmissing data on this endpoint, after applying the composite and treatment policy estimand, is less than 10 in the placebo arm in at least one stratum, NRI will be used instead of RBI.

a Liver-related clinical events include evidence of progression to decompensated cirrhosis, liver transplantation, or qualification for liver transplantation on basis of MELD score (Protocol Section 7.1.4.3)

To evaluate the overall effect of NASH resolution, Table 8 describes the estimand strategy for the secondary endpoint of NASH resolution without worsening in fibrosis (if applicable) at Week 72 in the SEMA+CILO/FIR versus placebo groups.

Table 8. Estimand Strategy for Secondary Endpoint 2: NASH Resolution without Worsening in Fibrosis at Week 72 in the SEMA+CILO/FIR Versus Placebo Groups

Secondary Endpoint 2	Hybrid Composite/Treatment Policy Strategy
Population	All evaluable participants (defined as participants with baseline re-read inflammation ≥ 1 and ballooning ≥ 1 or participants with liver-related clinical events or all-cause mortality prior to Week 72 liver biopsy) in the FAS as defined in Section 3.1.2
Patient-level outcomes to be measured	A binary response of NASH resolution (defined as lobular inflammation of 0 or 1 and hepatocellular ballooning of 0) without worsening in fibrosis (if applicable) at Week 72 in the SEMA+CILO/FIR vs placebo groups, based on the baseline re-read and postbaseline biopsy reading by the central pathologists.
Measure of intervention effect and handling of intercurrent events	Composite policy estimand: experienced a liver-related clinical event[a] or died due to any cause (all-cause mortality) prior to the Week 72 liver biopsy are assumed to be nonresponders for the secondary 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 72 liver biopsy (ie, dose interruption/reduction, protocol deviation, or premature study drug discontinuation).
Population level summary measure	Difference in proportions of participants with NASH resolution without worsening in fibrosis (if applicable) between SEMA+CILO/FIR and placebo groups.
Main estimators	A stratified Mantel-Haenszel test will be used to compare the difference in proportions of participants who meet NASH resolution without worsening in fibrosis (if applicable) at Week 72 between SEMA+CILO/FIR and placebo, adjusting for baseline diabetes status and ELF score (≥ 11.30 or < 11.30). Participants who experienced a liver-related clinical event or died due to any cause (all-cause mortality) prior to the Week 72 liver biopsy are considered nonresponders. Participants who had missing data at Week 72 due to reasons other than death or a liver-related clinical event[a] will be handled by reference-based imputation. If the number of participants with nonmissing data on this endpoint, after applying the composite and treatment policy estimand, is less than 10 in the placebo arm in at least one stratum, NRI will be used instead of RBI.

a Liver-related clinical events include evidence of progression to decompensated cirrhosis, liver transplantation, or qualification for liver transplantation on basis of MELD score (Section 7.1.4.3).

To demonstrate the contribution of SEMA to NASH resolution, Table 9 describes the estimand strategy for the secondary endpoint of NASH resolution without worsening in fibrosis (if applicable) at Week 72 in the SEMA+CILO/FIR versus CILO/FIR groups.

Table 9. Estimand Strategy for Secondary Endpoint 3: NASH Resolution without Worsening in Fibrosis at Week 72 in the SEMA+CILO/FIR versus CILO/FIR Groups

Secondary Endpoint 3	Hybrid Composite/Treatment Policy Strategy
Population	All evaluable participants (defined as participants with baseline re-read inflammation ≥ 1 and ballooning ≥ 1) in the FAS as defined in Section 3.1.2
Patient Level Outcomes to be Measured	A binary response of NASH resolution (defined as lobular inflammation of 0 or 1 and hepatocellular ballooning of 0) without worsening in fibrosis (if applicable) at Week 72 in the SEMA+CILO/FIR versus CILO/FIR groups, based on the baseline re-read and postbaseline biopsy reading by the central pathologists
Measure of intervention effect and handling of intercurrent events	Composite policy estimand: experienced a liver-related clinical event[a] or died due to any cause (all-cause mortality) prior to the Week 72 liver biopsy are assumed to be nonresponders for the secondary 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 72 liver biopsy (ie, dose interruption/reduction, protocol deviation, or premature study drug discontinuation).
Population level summary measure	Difference in proportions of participants with NASH resolution without worsening in fibrosis (if applicable) between the SEMA+CILO/FIR and CILO/FIR groups.
Main Estimators	A stratified Mantel-Haenszel test will be used to compare the difference in proportions of participants who meet NASH resolution without worsening in fibrosis at Week 72 between SEMA+CILO/FIR and CILO/FIR, adjusting for baseline diabetes status and ELF score (≥ 11.30 or < 11.30). Participants who experienced a liver-related clinical event or died due to any cause (all-cause mortality) prior to the Week 72 liver biopsy are considered nonresponders. Participants who had missing data at Week 72 due to reasons other than death or a liver-related clinical event[a] will be handled by reference-based imputation. If the number of participants with nonmissing data on this endpoint, after applying the composite and treatment policy estimand, is less than 10 in the placebo arm in at least one stratum, NRI will be used instead of RBI.

a Liver-related clinical events include evidence of progression to decompensated cirrhosis, liver transplantation, or qualification for liver transplantation on basis of MELD score (Protocol Section 7.1.4.3).

6.2.3. Analysis of the Secondary Histologic Endpoints

If the primary objective is achieved, the secondary efficacy endpoints will be tested sequentially as specified in Section 3.5.

A similar method of analysis as described for the primary analysis in Section 6.1.4 will be performed for the secondary endpoints. The point estimate, 95% CI and corresponding *P* value based on Z tests will be provided for the difference in proportions for each treatment comparison. The 95% CIs for proportion of responders in each treatment group (SEMA+CILO/FIR, SEMA, CILO/FIR, and Placebo).

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

6.2.4. Sensitivity Analysis of the Secondary Histologic Efficacy Endpoints

Nonresponder imputation (if not applied per the primary analysis estimand framework) and OC under the hybrid composite and treatment policy strategy of handling intercurrent events for sensitivity analysis (see details in Section 3.6.1) will be carried out for the secondary efficacy endpoints using the same method described in Section 6.2.3.

If the evaluable sample size after missing data imputation is less than 2 in a treatment group in a stratum, defined by baseline diabetes status and ELF score (≥ 11.30 or < 11.30), the stratification factor of baseline diabetes status will be excluded from the consideration. If the evaluable sample size is still less than 2 in a treatment group in a stratum, as defined by baseline ELF score, an unstratified version of the MH methods described in Section 6.1.4 will be applied.

CCI [redacted]

CCI

[redacted]

CCI

CCI [redacted]


CCI


CCI

6.5. Changes From Protocol-specified Efficacy Analysis

There are no deviations from the protocol-specified efficacy analyses, but clarifications were added to analyze data when described methods are not considered as appropriate under extreme cases. Specifically:

1) NASH resolution is updated as "NASH resolution without worsening in fibrosis" to incorporate potential cases with baseline re-read $\leq$ F3 fibrosis stage

2) Analysis of NASH resolution will only include evaluable participants, which are defined as participants with baseline re-read lobular inflammation and hepatocellular ballooning > 0, within the FAS.

3) For the primary analysis of the primary, secondary, and selected binary endpoints, RBI is replaced by NRI when the number of participants with nonmissing data on this endpoint, after applying the composite and treatment policy estimand, is less than 10 in the placebo arm in at least one stratum. For the stratified MH method, if the evaluable sample size after missing data imputation is less than 2 in a treatment group in a stratum based on baseline diabetes status and ELF score ($\geq$ 11.30 or < 11.30), the stratification factor of baseline diabetes status will be excluded from consideration. If the evaluable sample size is still less than 2 in a treatment group in a stratum based on baseline ELF score, an unstratified version of the Mantel-Haenszel (MH) methods described in Section 6.1.4 will be applied.

7. SAFETY ANALYSES

7.1. Adverse Events and Deaths

7.1.1. Adverse Event Dictionary

Clinical and laboratory adverse events (AEs) will be coded using the current version of 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 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. The missing category will be listed last in summary presentation. The by-participant data listings will show the missing severity grade as missing (blank).

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. Relationship of Adverse Events to Study Procedures

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

7.1.5. Relationship of Adverse Events to Pen-Injector

Study procedure related AEs are those for which the investigator selected "Related" on the AE case report form (CRF) to the question of "Related to Device (Pen-Injector)." Relatedness will always default to the investigator's choice, not that of the medical monitor. Events for which the investigator did not record relationships to device (pen-injector) will be considered related to pen-injector for summary purposes. However, by-participant data listings will show the relationship as missing.

7.1.6. 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 Patient Safety department before data finalization.

7.1.7. Treatment-Emergent Adverse Events

7.1.7.1. Definition of Treatment-Emergent Adverse Events

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 35 days after permanent discontinuation of study drug
- Any AEs leading to premature discontinuation of study drug.

7.1.7.2. Incomplete Dates

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 35 days after the date of the last dose of study drug

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.

7.1.7.3. Analysis of Crude Incidence Rate and Exposure-Adjusted Incidence Rates

The crude incidence rate (CIR) is defined as the percentage of participants with an event. The exact 95% CI {Clopper 1934} based on Binominal distribution will be used.

The exposure-adjusted incidence rate (EAIR) defined as below will be calculated for the TEAE tables as outlined in Section 7.1.8:

$$\text{Incidence rate per 100 patient-years of exposure (PYE)} = \frac{Total\ number\ of\ participants\ with\ an\ event}{Total\ PYE} \times 100$$

The total patient-years of exposure (PYE) to a treatment is the sum of the individual participant's PYE, which is defined as:

- For participants with an event:

$$PYE = (Event\ first\ start\ date - first\ dosing\ date + 1)/365.25$$

- When calculating PYE, a partial AE onset date will be imputed as follows:

  — If the non-missing portion of the AE onset date (month and/or year) is the same as the portion of the first dosing date (month and/or year) of start of study drug, the AE onset date will be imputed as the first dosing date of study drug; otherwise it will be imputed as the 1st day of month (if only day is missing) or January 1st of the year (if day and month are missing).

  — If the AE onset date is completely missing, the AE onset date will be imputed as the first dosing date of study drug.

- For participants with no event:

$$PYE = (Calculation\ end\ date - first\ dosing\ date + 1)/365.25,$$

where Calculation end date = min(date of the last dose of study drug + 35, study end date).

The exact 95% CI based on Poisson distribution {Ulm 1990} for EAIR is defined as:

$$\left(\frac{\chi^2_{2(Total\ number\ of\ participants\ with\ an\ event),0.025}}{2 \times Total\ PYE}, \frac{\chi^2_{2(1+Total\ number\ of\ participants\ with\ an\ event),0.975}}{2 \times Total\ PYE}\right) \times 100$$

Partial last dosing dates will be imputed as described in Section 4.2.1.

7.1.8. Summaries of Adverse Events and Deaths

Treatment-emergent AEs will be summarized based on the Safety Analysis Set.

7.1.8.1. Summaries of AE incidence

A brief, high-level summary of the number and percentage (CIR) and/or EAIR of participants who experienced at least 1 TEAE in the categories described below will be provided by treatment group. All deaths observed in the study will also be included in this summary.

The number and percentage (CIR) and/or EAIR of participants who experienced at least 1 TEAE will be provided and summarized by SOC, HLT, PT, and treatment group:

- TEAEs

For the AE categories described below, the number and percentage and/or EAIR will be provided by SOC, PT, and treatment group:

- TEAEs by maximum severity
- TEAEs with Grade 2 or higher
- TEAEs with Grade 3 or higher by maximum severity
- TEAE related to study procedure
- TEAE related to pen-injector
- TE treatment-related AEs by maximum severity
- TE treatment-related AEs with Grade 2 or higher
- TE treatment-related AEs with Grade 3 or higher by maximum severity
- TE SAEs
- TE treatment-related SAEs
- TEAE leading to premature discontinuation of study
- TEAEs leading to premature discontinuation of any study drug
- TEAEs leading to dose reduction or temporary interruption of any study drug
- TEAE leading to death
- TE gastrointestinal AEs, searched by SOC of gastrointestinal disorders
- TE pruritus-related AEs, utilizing a MST list developed by Gilead
- TEAE of special interest - hypoglycemic episodes according to the corresponding eCRF page
- TEAE of special interest - acute pancreatitis according to the corresponding eCRF page
- TEAE of special interest - acute gallbladder disease according to the corresponding eCRF page
- TEAE of special interest - malignant neoplasm (not including localized basal or squamous cell cancer or other localized non-melanoma skin cancer or carcinoma in situ of the cervix) according to the corresponding eCRF page

- TEAE of special interest - diabetic retinopathy according to the corresponding eCRF page
- TEAE of special interest - any case where a participant meets criteria for study drug withholding according to the corresponding eCRF page

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 total frequency within each SOC. 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, all TEAEs, TEAEs with Grade 3 or higher, TE SAEs, TE treatment-related AEs, TE treatment-related SAEs, TEAEs leading to premature discontinuation of any study drug, TEAEs leading to dose reduction or temporary interruption, TEAEs leading to premature discontinuation of study, TEAE related to pen-injector and TEAE related to study procedure will be summarized using number and percentage by PT only, in descending order of total frequency.

In addition, separate summary table of all AEs from the date of first dose to the end of study and summary table of all AEs that occurs after the last dose + 35 days up to the end of the study will be provided. The number and percentage of participants who experienced at least 1 such AE will be provided and summarized by SOC, PT, and treatment group.

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 with severity of Grade 2 or higher
- All AEs leading to premature discontinuation of study drug
- All AEs leading to temporary interruption of any study drug
- All AEs leading to dose reduction of injectable study drug
- All AEs leading to premature discontinuation of study
- All AESI (including hypoglycemic episodes, acute pancreatitis, acute gallbladder disease, malignant neoplasm (not including localized basal or squamous cell cancer or other localized non-melanoma skin cancer or carcinoma in situ of the cervix), diabetic retinopathy, and any case where a participant meets criteria for study drug withholding)

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 the Safety Analysis Set and will include data collected up to the last dose of study drug plus 35 days for participants who have permanently discontinued study drug, or all available data at the time of the database snapshot for participants who were ongoing at the time of an interim analysis (eg, DMC). 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.

A by-participant listing for laboratory test results will be provided by participant ID number and visit in chronological order for hematology and serum chemistry separately. Values falling outside of the relevant reference range and/or having a severity grade of 1 or higher on the CTCAE severity grade (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 treatment group for creatinine, white blood cells (WBC), neutrophils, lymphocytes, hematocrit, and hemoglobin as follows:

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

A baseline laboratory value will be defined as the last measurement obtained on or prior to the date of first dose of study drug. 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.

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 the baseline ALP, total bilirubin, direct bilirubin, GGT, ALT, AST, and LDL toxicity grades, the CTCAE version 5.0 will be used to assign grades to the derived average values.

7.2.2.1. Treatment-Emergent Laboratory Abnormalities

Treatment-emergent laboratory abnormalities (TELAs) 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 35 days. 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

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 35 days. If the relevant baseline laboratory value is missing, any Grade 3 or 4 values observed within the timeframe specified above will be considered TE marked abnormalities.

7.2.2.3. Summaries of Laboratory Abnormalities

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

The following summaries (number and percentage of participants) for TELAs will be provided by lab test and treatment group; 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, the denominator is the number of participants with nonmissing postbaseline values up to 35 days after last dosing date for participants who have permanently discontinued study drug, or all available data at the time of the database snapshot for participants who were ongoing at the time of an interim analysis (eg, for DMC).

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

7.2.3. Liver-related Laboratory Evaluations

The summary will be provided by 2 subgroups of participants according to their baseline ALT/AST/ALP level:

- Normal
- Elevated (> upper limit of normal [ULN])

Liver-related abnormalities after initial study drug dosing will be examined and summarized using the number and percentage of participants who were reported to have the following laboratory test values for postbaseline measurements:

For participants with normal baseline ALT/AST/ALP (evaluated independently):

- Participants meeting criteria for close observation (any of the following):
  - ALT/AST/ALP > 3 × ULN
- Participants meeting criteria for study drug withheld (any of the following):
  - ALT/AST > 8 × ULN
  - ALT/AST > 5 × ULN for 2 weeks
  - ALT/AST/ALP > 3 × ULN and total bilirubin > 2 × ULN (or direct bilirubin > 2 × baseline in participants with Gilbert's syndrome or other cause of unconjugated hyperbilirubinemia)
  - ALT/AST/ALP > 3 × ULN and INR > 1.5 (applicable only to participants not on therapeutic anticoagulation)

For participants with baseline ALT/AST/ALP > ULN:

- Participants meeting criteria for close observation (any of the following):
  - ALT/AST/ALP > 2 × Baseline
  - ALT/AST > 300 U/L
- Participants meeting criteria for study drug withheld (any of the following):
  - ALT/AST > 500 U/L
  - ALT/AST > 3 × Baseline and total bilirubin > 2 × ULN (or direct bilirubin > 2 × Baseline in participants with Gilbert's syndrome or other cause of unconjugated hyperbilirubinemia)
  - ALT/AST > 300 U/L, and total bilirubin > 2 × ULN (or direct bilirubin > 2 × Baseline in participants with Gilbert's syndrome or other cause of unconjugated hyperbilirubinemia)

- ALT/AST > 3 × Baseline and INR > 1.5 (if not on therapeutic anticoagulation)
- ALT/AST > 300 U/L and INR > 1.5 (if not on therapeutic anticoagulation)
- ALP > 2 × Baseline and total bilirubin > 2 × ULN (or direct bilirubin > 2 × Baseline in participants with Gilbert's syndrome or other cause of unconjugated hyperbilirubinemia)
- ALP > 2 × Baseline and INR > 1.5 (if not on therapeutic anticoagulation)

The summary will include data from all postbaseline visits up to 35 days after the last dose of study drug for participants who have permanently discontinued study drug, or all available data at the time of the database snapshot for participants who were ongoing at the time of an interim analysis (eg, for DMC). For individual laboratory tests, participants will be counted once based on the most severe postbaseline values. For the composite criteria of AST, ALT, or ALP and total bilirubin, direct bilirubin, or INR, participants will be counted once when the criteria are met at the same postbaseline visit date. The denominator is the number of participants in the Safety Analysis Set who have nonmissing postbaseline values of all relevant tests at the same postbaseline visit date.

Separate summary tables for close observation and drug withholding confirmed by repeated testing of laboratory test(s) will also be provided. The repeated testing is defined as the laboratory test(s) immediately after the visit when close observation or drug withholding was initially flagged (evaluated independently for each criterion as above). For the composite criteria of AST, ALT or ALP and total bilirubin, direct bilirubin or INR, the laboratory test(s) from the same postbaseline visit date immediately after the visit of triggering criteria will be considered.

A listing of participants who met at least 1 of the above criteria for close observation and drug withholding will be provided separately.

7.3. Body Weight, Anthropometric Parameters and Vital Signs

Descriptive statistics will be provided by treatment group for waist and hip circumference, BMI and vital signs (pulse [bpm], respiratory rate [breath/min], and temperature [°C]) as follows:

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

A baseline value will be defined as the last available value collected on or prior to the date of first dose of study drug. Change from baseline to a postbaseline visit will be defined as the postbaseline value minus the baseline value. Descriptive statistics for body weight and blood pressure are included in Section 6.4.

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, height, BMI, hip and waist circumference 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 as of the final analysis.

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 for each treatment group. 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 overall frequency. 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. 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.

For the purpose of computing the study end date, the partial medication start and/or stop dates will be imputed as follows:

- For a medication with a partial start date, if the month and year are available, the missing day will be imputed as 01; if only year is available, both day and month will be imputed as 01 January.

- For a medication with a partial stop date, if the month and year are available, the missing day will be imputed as the last day of the month; if only year is available, both day and month will be imputed as 31 December.

7.5. Electrocardiogram Results

7.5.1. Investigator Electrocardiogram Assessment

A shift table of the investigators' assessment of ECG results at each visit compared with baseline values will be presented by treatment group using the following categories: normal; abnormal, not clinically significant; abnormal, clinically significant; or missing. The number and percentage of participants in each cross-classification group of the shift table will be presented. Participants with a missing value at baseline or postbaseline will not be included in the denominator for percentage calculation. No formal statistical testing is planned.

A by-participant listing for ECG assessment results will be provided by participant ID number and visit in chronological order.

7.6. Other Safety Measures

No additional safety measures are specified in the protocol.

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

Single PK plasma samples will be collected and archived for PK analysis of CILO and FIR (and their metabolites, as applicable). Samples will be collected at Week 4 (15 minutes to 3 hours postdose), Week 24 (anytime postdose), Week 48 (predose), Week 60 (15 minutes to 3 hours postdose), Week 72 (predose), the ET visit (anytime), and at unscheduled visits (anytime) that are performed for the purpose of safety evaluation (eg, SAE follow-up). For PK sampling at Weeks 4, 48, 60, and 72, participants should be reminded not to take their oral study drug until advised to do so at their clinic visit.

8.2. PK Analyses Related to Sparse PK Sampling

A population PK model may be developed to characterize the PK of CILO, FIR, and their respective metabolite(s), if applicable. Sparse PK sampling data from this study may be combined with data from other studies in a metapopulation analysis using mixed-effect modeling techniques. Details of the population PK analysis will be provided in a separate PK analysis plan, if applicable.

The following listing will be provided for all PK samples collected in this study:

- PK sampling details by participant including actual dosing time and actual draw time, calculated time postdose of sample collection, sample age, and sample concentration by analyte.

8.3. Changes From Protocol-Specified PK Analyses

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

9. REFERENCES

Clopper CJ, Pearson ES. The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. Biometrika 1934;26 (4):404-13.

Doward LC, Balp M-M, Twiss J, Slota C, Cryer D, Langford A, et al. Measuring What Matters to Patients: The Development of the Nash Check, A New Patient Reported Outcome Instrument for Non Alcoholic Steatohepatitis [Poster]. European Association for the Study of the Liver (EASL); 2018 April 11-15; Paris, France.

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.

Ulm K. A simple method to calculate the confidence interval of a standardized mortality ratio (SMR). Am J Epidemiol 1990;131 (2):373-5.

10. SOFTWARE

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

nQuery Advisor(R) Version 4.0. Statistical Solutions, Cork, Ireland.

11. SAP REVISION

Revision Date (DD MMM YYYY)	Section	Summary of Revision	Reason for Revision
Final Draft v1.0 (30 July 2021)		Original version	
Final Draft v2.0 (01 Mar 2024)	1.1, 1.3, 3.5, 6.1.2, 6.2.2	Update objective, sample size, power, multiple comparison, estimands for primary and secondary endpoints	Align with protocol amendment 2
	3.4	Add additional subgroup analysis	To take into consideration of variation of baseline biopsy re-read
	3.6, 6.1.5, 6.2.4, 6.3, 6.4, 7.1.7.3	• Remove MI and TPA for missing data imputation in sensitivity analysis • Remove a few sensitivity analyses. • CCI [redacted] • Remove treatment comparison for EAIR	To simplify statistical analysis to support objectives of this Ph2 study
	3.6, 4.2.2, 7.1.4, 7.1.5, 7.2.3, 7.4.2	• Clarify analysis and missing data imputation method under scenario where one stratum is too small for analyses • Add an additional calculation method to injectable drug adherence • Clarify approach to handle missing of relatedness to study procedure or pen-injector • Add additional tables for liver-related close observation and drug withholding considering repeated testing • Clarify approach to handle partial dates in concomitant medication	Provide clarification for analysis
	7.1.8	• Add EAIR and confidence interval for EAIR and CIR for TEAE tables • Add additional summary table for AEs that occurred after last dose + 35 days up to study end	To align with FDA request

Revision Date (DD MMM YYYY)	Section	Summary of Revision	Reason for Revision
Final v1.0 (07 JAN 2025)	3.6	• Add MI and TPA for primary efficacy endpoint	To align with FDA request
	3.8.3, 5.2, 6.1, 6.2, 6.3, 6.4, 6.5, 7.1.8.1	• Add details for selection of multiple records for biopsy data • Add "without worsening in fibrosis" to efficacy endpoints of NASH resolution to incorporate potential cases with baseline re-read F3 firbosis stage • Add more details for baseline characteristics summary and efficacy analyses • Add change from protocol-specified efficacy analyses • Change to separate tables for each category of AE of interest	Provide clarification for analysis
	4.4, 6.4	• Remove repeated analysis for COVID-19 impact • Remove repeated endpoints which are included in section 6.1-6.3	Remove repeated information
	7.1.8	• Add the summary table for "all AEs from the date of first dose to the end of study"	To align with FDA request

12. APPENDICES

Appendix 1. Schedule of Assessments

	Screening	Day 1	Day 29 (Wk 4) (± 3 d)	Day 57 (Wk 8) (± 3 d)	Day 85 (Wk 12) (± 3 d)	Day 113 (Wk 16) (± 3 d)	Day 169 (Wk 24) (± 3 d)	Day 253 (Wk 36) (± 3 d)	Day 337 (Wk 48) (± 3 d)	Day 421 (Wk 60) (± 3 d)	Day 505 (Wk 72) (−14 d)	ET	Follow-up (± 7 d)[a]
Written informed consent	X												
Confirm eligibility	X	X											
Medical history	X												
Symptom-driven physical examination	X[b]	X	X	X	X	X	X	X	X[b]	X	X[b]	X[b]	X
Assessment of ascites and hepatic encephalopathy[c]	X	X	X	X	X	X	X	X	X	X	X	X	X
Vital signs including body weight	X	X	X	X	X	X	X	X	X	X	X	X	X
Height	X												
Hip and waist circumference	X	X					X		X		X	X	X
Fundus examination	X[d]								X[d]		X[d]	X[d,e]	
12-lead ECG[f]	X										X	X[e]	
Calculation of the CP and MELD scores	X	X	X	X	X	X	X	X	X	X	X	X	X
Liver biopsy[g]	X										X	X[e]	
Abdominal ultrasound	X						X		X		X	X[h]	
FibroScan®	X						X		X		X	X[h]	
PRO measures		X[i]					X		X		X	X[e]	
Lifestyle counseling	X	X	X	X	X	X	X	X	X	X	X	X	X
Adverse events[j]	X	X	X	X	X	X	X	X	X	X	X	X	X
Concomitant medications	X	X	X	X	X	X	X	X	X	X	X	X	X
Dispense study drug		X	X	X	X	X	X	X	X	X			
Review of study drug dosing compliance			X	X	X	X	X	X	X	X	X	X	
Counseling regarding adherence to study procedures		X	X	X	X	X	X	X	X	X	X		
Participant fasting	X	X	X	X	X	X	X	X	X	X	X	X	X
Chemistry, hematology, and coagulation panels[k]	X	X	X	X	X	X	X	X	X	X	X	X	X

	Screening	Day 1	Day 29 (Wk 4) (± 3 d)	Day 57 (Wk 8) (± 3 d)	Day 85 (Wk 12) (± 3 d)	Day 113 (Wk 16) (± 3 d)	Day 169 (Wk 24) (± 3 d)	Day 253 (Wk 36) (± 3 d)	Day 337 (Wk 48) (± 3 d)	Day 421 (Wk 60) (± 3 d)	Day 505 (Wk 72) (−14 d)	ET	Follow-up (± 7 d)[a]
Glycemic panel	X	X			X		X		X		X	X	X
Lipid panel	X	X	X	X	X	X	X	X	X	X	X	X	X
HbA_{1c}	X	X			X		X		X		X	X	X
eGFR by MDRD	X	X	X	X	X	X	X	X	X	X	X	X	X
ELF test	X	X			X		X		X		X	X[e]	
SARS-CoV-2 RT-PCR Test[l]	X												
Pregnancy testing[m]	X	X	X	X	X	X	X	X	X	X	X	X	X
Follicle-stimulating hormone[n]	X												
Single PK sampling[o]			X				X		X	X	X	X	
Blood collection (biomarkers)	X	X			X		X		X		X	X[e]	X
Urine collection (biomarkers)	X	X			X		X		X		X	X[e]	X
Stool collection (biomarkers)[p]		X							X		X		
Urine drug screen	X												
HIV-1, HBV, HCV serology	X												

[redacted]

CK = creatinine kinase; COVID-19 = coronavirus disease 2019; CP = Child-Pugh; CTCAE = Common Terminology Criteria for Adverse Events; d = days; ECG = electrocardiogram; eGFR = estimated glomerular filtration rate; ELF = enhanced liver fibrosis; ET = early termination; FSH = follicle-stimulating hormone; HBV = hepatitis B virus; HCV = hepatitis C virus; HbA_{1c} = hemoglobin A_{1c}; HIV-1 = human immunodeficiency virus type 1; MDRD = Modification of Diet in Renal Disease equation; MELD = Model for End-Stage Liver Disease; PK = pharmacokinetic(s); PGIC = Patient Global Impression of Change; PRO = patient-reported outcome; RT-PCR = reverse transcriptase-polymerase chain reaction; SAE = serious adverse event; SARS-CoV-2 = severe acute respiratory syndrome coronavirus 2; Wk = week

a Participants will complete a follow-up visit 5 weeks after the Week 72 or ET visit.

b Perform complete physical examination at screening visit only. See Appendix 11.9.1 for France-specific text.

c Assessment is to be done in accordance with established standard practices at the site and if present, graded based on CTCAE or West Haven criteria (Appendix 11.4), respectively.

d For participants with type 2 diabetes diagnosed prior to the date of the screening visit OR based on screening visit $HbA_{1c} \geq 6.5\%$, participants must have no evidence of uncontrolled and potentially unstable retinopathy or maculopathy as determined by a fundoscopic examination performed starting 90 days prior to screening visit date through Day 1. If there has been worsening of the participant's visual function since a historical fundoscopic examination in the opinion of the investigator, then the fundoscopic examination must be repeated prior to Day 1 for eligibility. Pharmacological pupil dilation is a requirement unless using a digital fundus photography camera specified for nondilated examination.

e Assessments to be performed at the discretion of the investigator.
f See Appendix 11.9.1 for France specific text.
g Required for all participants at time of screening unless a qualifying liver biopsy is available per inclusion criterion 3. On-treatment liver biopsy required at Week 72. Early termination liver biopsy at the discretion of the investigator. CCI [redacted]
h Assessment to be performed at the ET visit, unless performed within 12 weeks prior to the ET visit.
i PGIC Pain and Fatigue will not be assessed at Day 1.
j Adverse events reporting during screening is limited to SAEs and adverse events related to study procedures.
k If a participant is receiving treatment with both a fibrate and a statin, serum CK should be measured at study visits and appropriate clinical evaluation, including interruption of statin therapy, should be initiated if muscle injury is suspected.
l For participants who have not completed a series of an authorized COVID-19 vaccination regimen prior to screening.
m All women of childbearing potential will have a serum pregnancy test at screening. Urine pregnancy testing will occur at Day 1 (prior to dosing) and on each treatment visit. Starting at the Week 4 visit, every 4 weeks (± 3 days) urine pregnancy testing may be performed at home in between clinic visits, where possible, or in clinic, if at home pregnancy tests are unavailable.
n Serum FSH test (only for women who are < 54 years old and have stopped menstruating for ≥ 12 months but do not have documentation of ovarian hormonal failure)
o Single PK sample at Week 4 (15 minutes to 3 hours postdose), Week 24 (anytime postdose), Week 48 (predose), Week 60 (15 minutes to 3 hours postdose), Week 72 (predose), and ET visit (anytime). In cases where the unscheduled visit is performed for the primary purpose of safety evaluation (eg, SAE follow-up), collection of a single PK sample (anytime predose or postdose) is recommended. The timing of the PK sample in relation to the last dose of study drug should be documented. For PK sampling at Weeks 4, 48, 60, and 72, participants should be reminded not to take their oral study drug until advised to do so at their clinic visit.
p Participants will be given a stool sample collection kit at the screening visit and Weeks 36 and 60. Stool sample collection should be completed by the participant in advance of the required study visits (Day 1 and Weeks 48 and 72).
q CCI [redacted]

Appendix 2. CTCAE Grade for Laboratory Parameters

Common Terminology Criteria for Adverse Events (CTCAE) Version 5.0, https://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_5.0

Appendix 3. Liver Function Prognostic Scores

Child-Pugh Score Calculation

	1	2	3
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** Abdominal distension Medication for ascites	**Severe (diuretic-refractory)** Visible clinically
Total 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

HE = hepatic encephalopathy; INR = international normalized ratio

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 (or same day if same blood draw is not available). 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 (or same day if same blood draw is not available) 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).

MELD Score Calculation

- 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 the participant has had 2 or more dialysis treatments within the preceding week. 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.

Appendix 4. Patient-reported Outcome Measures (PROs)

It is recommended that these questionnaires, as possible, be completed prior to any other study assessments at each visit. The participant should read the questionnaires and record the answers independently.

1. FACIT-Fatigue

The FACIT-Fatigue PRO measure is a 13-item questionnaire designed to measure fatigue and lifestyle-related consequences of fatigue, with all responses reported on a scale of 0 to 4. Participants are directed to respond to each question based on their experience in the previous 7 days.

If more than 50% items (ie, 7 out of 13) were answered, the Fatigue Subscale Score is defined as (sum of non-missing scores)*13/number of non-missing items. Otherwise, the fatigue subscale score will be considered as missing.

2. NASH-CHECK

The NASH-CHECK is a 28-item PRO measure designed to measure symptoms and health-related quality of life including day-to-day activities, emotions, and lifestyle in participants with NASH. Symptoms of NASH are reported on a scale of 0 to 10, with participants asked to report the severity of the symptom at its worst over the previous 7 days. Participants are directed to respond to additional questions related to day-to-day activities, emotions and lifestyle based on quality of life in the prior 7 days. The 28 items are used to derive six symptom scale scores (abdominal pain, abdominal bloating, fatigue, sleep, itchy skin, and cognitive symptoms) and three HRQOL scale scores (activity limitations, emotional impact, and social impact). Details of the derivation of symptom scale scores and HRQOL scale scores refer to NASH-CHECK scoring guideline {Doward 2018}

3. Patient Global Impression Scale of Severity (PGI-S) Fatigue

For the PGI-S Fatigue, participants will be asked to rate their overall impression of their fatigue during the past week on a four-point scale from “none” to “severe”. The PGI-S Fatigue will be used as an anchor to support the interpretation of the FACIT-Fatigue questionnaire and NASH-CHECK fatigue items.

4. Patient Global Impression Scale of Change (PGI-C) Fatigue

For the PGI-C Fatigue, participants will be asked to rate their overall impression of how their fatigue has changed since they started taking the study medication, on a seven-point scale from “very much improved” to “very much worsened”. The PGI-C Fatigue will be used as an anchor to support the interpretation of the FACIT-Fatigue and NASH-CHECK fatigue items.

5. PGI-S Pain

For the PGI-S Pain, participants will be asked to rate their overall impression of their upper abdominal (stomach) pain during the past week on a four-point scale from “none” to “severe”. The PGI-S Pain will be used as an anchor to support the interpretation of the NASH-CHECK upper abdominal (stomach) pain item.

6. PGI-C Pain

For the PGI-C Pain, participants will be asked to rate their overall impression of how their upper abdominal (stomach) pain has changed since they started taking the study medication, on a seven-point scale from “very much improved” to “very much worsened”. The PGI-C Pain will be used as an anchor to support the interpretation of the NASH-CHECK upper abdominal (stomach) pain item.

7. SF-36

The SF-36 consists of 36 questions to measure functional health and well-being from the participant’s point of view, and comprises 8 health domains (physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, and mental health). These health domain scales contribute to the physical health and mental health summary measures.

Appendix 5. NAFLD Activity Score (NAS)

NAS score is calculated as the sum of steatosis grade, lobular inflammation grade and hepatocellular ballooning grade. If any of its elements are missing, then NAS will be also missing.

Appendix 6. Noninvasive markers for Fibrosis

- **Enhanced Liver Fibrosis (ELF™) Score**

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

Note: All ELF™ score components (hyaluronic acid (ng/mL), PIIINP (ng/mL) and TIMP1 (ng/mL)) need to be measured from the same blood draw. If any component is less than the lower LOQ or above the upper LOQ, the component will be imputed per data handling conventions in Section 3.7, and ELF™ score will be calculated based on the imputed values of components.

- **Fibrosis-4 (FIB-4) Index Calculation:**

FIB-4 Index = round ((age (years) × AST (U/L)) / (platelet ($\times 10^9$/L) × sqrt (ALT (U/L))), 0.01);

- **AST to Platelet Ratio Index (APRI) Calculation:**

APRI = round (AST (U/L)/ASTULN (U/L) × 100 / platelet ($\times 10^9$/L), 0.1);

Note: ASTULN is Upper limit of normal in AST. For FIB-4 index and APRI calculation, the laboratory parameters need to be measured from the same blood draw. Age should be the actual age at the date when laboratory values are taken.

- **NAFLD fibrosis score (NFS) Calculation:**

NFS = -1.675 + 0.037 × age (years) + 0.094 × BMI (kg/m^2) + 1.13 × IFG (impaired fasting glucose) / pre-diabetes or diabetes (yes = 1, no = 0) + 0.99 × AST / ALT ratio - 0.013 × platelet ($\times 10^9$/L) - 0.66 × albumin (g/dL). Keep 3 decimal places.

The laboratory parameters need to be measured from the same blood draw. The last non-missing BMI on or prior to laboratory date should be used. Age should be the actual age at the date when laboratory values are taken. Status of pre-diabetes/diabetes should also be decided on the laboratory date. If a participant had pre-diabetes/diabetes at baseline, the pre-diabetic/diabetic status will be yes for all the postbaseline visits. If the participant does not have pre-diabetic/diabetes at baseline, the diabetic status will be determined based on the start date AEs of pre-diabetes and diabetes, and the collection date when fasting glucose is greater than 100 (IFG). If the AE start date or the fasting glucose collection date is on or prior to the laboratory date of a specific visit, then the diabetic/pre-diabetic status will be yes for that visit and later visits. If the day of the AE start date is missing, it will be imputed using the 1st day of the month.

- **FibroSURE/FibroTest**

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

Note: 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.

Appendix 7. Data Collection and Determination of Disaster or Public Health Emergency Data

This appendix describes the clinical trial site collection of Disaster or Public Health Emergency data and the data processing algorithm that will be used to determine Disaster or Public Health Emergency related information from CRF comment fields.

12.1 Data Collection

Missed or virtual visits due to Disaster or Public Health Emergency will be collected through the visit date CRF page. However, some Disaster or Public Health Emergency related information will be collected via CRF comment fields (eg. comments related to study drug or study discontinuation due to Disaster or Public Health Emergency, missed key assessment due to Disaster or Public Health Emergency).

12.2 Determination of Disaster or Public Health Emergency Related Information from CRF Comment

Natural Language Processing (NLP) will be used to search the CRF comment fields to identify instances of "COVID-19", "Virtual", or synonyms (see Table 10 for example for COVID-19 pandemic). The search terms will be maintained in a global lookup table and can be modified to tune the NLP model. If COVID-19 and Virtual terms are identified through NLP, it will help the determination of study drug or study discontinuation due to Disaster or Public Health Emergency for certain reasons (eg. participant decision, withdrew consent, or investigator discretion), and provide supplemental information to determine missed key assessments due to Disaster or Public Health Emergency.

Table 10. Example Search Terms for "COVID-19" and "Virtual"

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-454-6075-SAP-Final-v1.0

ELECTRONIC SIGNATURES

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:mm:ss)
PPD	Biostatistics eSigned	10-Jan-2025 06:38:24
PPD	DRC Chair eSigned	10-Jan-2025 21:42:13