1. Title Page

STATISTICAL ANALYSIS PLAN

AURORA: A Phase 3, Multicenter, Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Efficacy and Safety of Cenicriviroc for the Treatment of Liver Fibrosis in Adult Subjects with Nonalcoholic Steatohepatitis

Protocol Number:	3152-301-002
Development Phase:	3
Product Name:	Cenicriviroc mesylate (CVC)
Study Statistician:	Grace West
Sponsor:	Allergan plc
Amendment 1 Date:	October 17, 2018
Amendment 2 Date:	February 19, 2020

2. Table of Contents

2.1 List of Tables

2.2 List of Figures

No table of figures entries found.

3. List of Abbreviations and Definition of Terms

Table 3-1 Abbreviations and Definitions of Terms

Abbreviation/Term	Definition
AE	adverse event
AESI	adverse event(s) of special interest
ALP	alkaline phosphatase
ALT	alanine aminotransferase
APRI	aspartate aminotransferase to platelet count ratio index
AST	aspartate aminotransferase
ANCOVA	analysis of covariance
ATC	Anatomical Therapeutic Chemical
BMI	body mass index
CFB	change from baseline
CI	confidence interval
CMH	Cochran-Mantel-Haenszel
CRN	Clinical Research Network
CVC	cenicriviroc mesylate
DSMB	Data and Safety Monitoring Board
DILI	drug-induced liver injury
eCRF	electronic case report form
ECG	Electrocardiogram
ELF	Enhanced Liver Fibrosis
FIB-4	noninvasive hepatic fibrosis index score combining standard biochemical values (platelets, ALT, AST) and age
FWER	familywise error rate
GLP-1	glucagon-like peptide 1
HbA1c	hemoglobin A1c
HBcAb	hepatitis B core antibody
HBsAb	hepatitis B surface antibody
HBsAg	hepatitis B surface antigen
HCV	hepatitis C virus
HCVAb	hepatitis C virus antibody
HR	Heart rate
HR	hazard ratio
IgG	immunoglobulin G
IL	Interleukin
INR	international normalized ratio
ITT	intent-to-treat
LDH	lactate dehydrogenase
LDL	low-density lipoprotein
LS	least squares
MACE	major adverse cardiovascular events
MAR	Missing at random
MedDRA	Medication Dictionary for Regulatory Activities
MELD	Model of end stage liver disease
MNAR	Missing not at random

Abbreviation/Term	Definition
mITT	modified intent-to-treat
NAS	NAFLD activity score
NASH	nonalcoholic steatohepatitis
NCI CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
NFS	NAFLD fibrosis score
OR	odds ratio
PCS	potentially clinically significant
PD	pharmacodynamics(s)
PK	pharmacokinetic(s)
PK/PD	pharmacokinetic/pharmacodynamic
PID	participant identification
PMM	pattern-mixture model
PP	per-protocol
[redacted]	
PT	preferred term
QT	time between the start of the Q wave and the end of the T wave (ECG)
QTc	QT interval corrected for heart rate
QTcB	QT interval corrected for heart rate using the Bazett formula ($QTcB = QT/(RR)^{1/2}$)
QTcF	QT interval corrected for heart rate using the Fridericia formula ($QTcF = QT/(RR)^{1/3}$)
SAE	serious adverse event
SAP	statistical analysis plan
[redacted]	
SI	Le Système International d'Unités (International System of Units)
SMQ	Standard MedDRA query
SOC	system organ class
T2DM	type 2 diabetes mellitus
TE	transient elastography
TEAE	treatment-emergent adverse event
TTE	Time-to-event
WHO	World Health Organization
[redacted]	

4. Introduction

This statistical analysis plan (SAP) details comprehensive, technical specifications of the statistical analyses of the efficacy and safety data outlined and/or specified in the final protocol of Study 3152-301-002 and the most recent amendment (version 4) dated 25 April 2019, along with the region-specific protocol amendment for the European Union (EU-1) dated 06 May 2019. Specifications of tables, figures, and data listings are contained in a separate document. The SAPs for pharmacokinetic/pharmacodynamic data and for health economics and outcomes research data will be prepared separately.

This document is organized into 3 main sections:

1. Study Design Summary
2. Statistical Methodology and Study Endpoints
3. Data Handling and Analysis Conventions

4.1 Study Design Summary

4.1.1 Overall Design

This is a Phase 3, multicenter, randomized, double-blind, placebo-controlled study of CVC for the treatment of Stage 2 or 3 liver fibrosis in adult subjects with NASH. The study will be conducted in 2 parts: Part 1 (histopathological surrogate endpoint) and Part 2 (clinical outcomes).

In Part 1, approximately 1200 subjects will be randomized 2:1 to receive CVC or placebo to assess the effect of CVC on the histopathological surrogate endpoint at Month 12. Subjects continuing beyond Month 12 will remain in the same treatment arm throughout the end of the study.

Part 2 of the study will focus on assessing the effect of CVC on clinical outcomes and will include approximately 2000 subjects, of which approximately 1200 subjects will have been randomized in Part 1.

The study population will include adult subjects with a liver biopsy diagnosis of NASH and Stage 2 or 3 liver fibrosis (by NASH CRN system) as confirmed by an independent central pathologist. CVC and placebo will be administered as double-blinded study drug. Study drug (CVC or placebo) must be taken once daily with food. A Data and Safety Monitoring Board (DSMB) will be formed to review ongoing data from Part 1 and Part 2 of this study.

The study will be terminated when adjudicated events have been accrued in approximately 367 unique subjects across Part 1 and Part 2. All subjects will have an End of Study visit at that time and will return to the clinic for a Follow-up visit 30 days after the last dose of study drug.

Subjects who permanently discontinue study drug before planned end of study are encouraged to complete all subsequent study visits and participate (without treatment) in Part 2 of the study for assessment of clinical outcomes (at the discretion of the investigator, subjects who permanently discontinue study drug may be followed every 6 months unless otherwise specified by the protocol). Subjects who undergo early termination from the study will return to the clinic within 48 hours for an early discontinuation visit and will return to the clinic for a Follow-up visit 30 days after the last dose of study drug.

Survival data will be collected for all study subjects at the end of the study.

Subjects who progress to cirrhosis or reach an adjudicated liver-related clinical outcome will complete the study and return to the clinic for a Follow-up visit (Visit 98) 30 days after the last dose of study drug. After completion of the Follow-up visit (Visit 98), subjects will become eligible for open-label access to CVC in a separate study (Clinical Study Protocol 3152-201-002).

Part 1

At Baseline (Day 1), eligible subjects will be assigned to the treatment arms using permuted block randomization stratified by NASH CRN fibrosis stage (2 or 3) and presence or absence of documented type 2 diabetes mellitus (T2DM) (yes or no). Eligible subjects will be randomized 2:1 to one of the following 2 treatment arms:

Arm	N	Treatment
A	800	CVC 150 mg, once daily
B	400	Placebo, once daily

All subjects will undergo safety assessments at the Baseline visit and at Months 1, 3, 6, 9, and 12. The liver biopsy, used to evaluate the efficacy endpoints, must be performed within ± 2 weeks of the Month 12 visit (Visit 7).

Part 2

In Part 2, Baseline (Day 1) eligible subjects will be randomized to treatment using permuted block randomization, stratified by presence or absence of documented T2DM (yes or no). Newly enrolled subjects will be randomized 2:1 to one of the following 2 treatment arms to achieve a total of approximately 2000 randomized subjects in Part 2 (of which approximately 1200 subjects will have been randomized in Part 1):

Arm	Newly Randomized Subjects in Part 2 N	Total Subjects (Randomized in Part 1 or Part 2) N	Treatment
A	534	1334	CVC 150 mg, once daily
B	266	666	Placebo, once daily

Subjects who enter Part 2 from Part 1 of the study will have study visits every 3 months until Part 2 is completed. Study drug will be dispensed and adverse events (AEs), concomitant medications, and adherence to study drug will be reviewed at every visit (every 3 months). Safety laboratory assessments will be performed every 6 months through the end of the study.

Subjects who are randomized in Part 2 and who did not participate in Part 1 of the study will undergo safety laboratory assessments at Months 1, 6, and 12 and every 6 months thereafter through the end of the study. Beginning at Month 3, AEs, concomitant medications, and adherence will be reviewed every 3 months through the end of the study.

Newly randomized subjects in Part 2 will undergo a liver biopsy at Screening for assessment of eligibility and a further liver biopsy at Month 12 (within ± 2 weeks). All subjects in Part 2 will undergo a liver biopsy at Month 60 (within ± 2 weeks) after the first dose of study drug.

4.1.2 Number of Participants

Part 1:

Approximately 1200 subjects will be randomized in Part 1; approximately 800 subjects are to receive CVC and approximately 400 subjects are to receive placebo. When the approximately 1200 subjects in Part 1 have also completed their 12-month visit assessments (including biopsy), the Part 1 analysis will commence. Subjects who remain in Part 1 at Month 12 will continue to receive the same treatment assignment in Part 2.

Part 2:

Approximately 2000 subjects (of which approximately 1200 subjects will have been randomized in Part 1); approximately 1334 subjects are to receive CVC and approximately 666 subjects are to receive placebo.

4.2 Study Objectives and Endpoints

Each global study objective is presented with corresponding endpoint(s) below.

Global Part 1:

Objectives	Endpoints
Primary • To demonstrate the superiority of CVC compared to placebo on liver histology at Month 12 relative to the Screening biopsy	**Primary Endpoint** • Improvement in fibrosis by at least 1 stage (NASH CRN system) AND no worsening of steatohepatitis (no worsening of lobular inflammation or hepatocellular ballooning grade as per 2005 criteria) on liver histology at Month 12 relative to Screening biopsy
Key Secondary • Evaluate the effect of CVC compared to placebo on liver histology at Month 12 relative to the screening biopsy for the proportion of subjects with improvement in fibrosis of at least 2 stages (NASH CRN system) AND no worsening of steatohepatitis (no worsening of lobular inflammation or hepatocellular ballooning grade 2005)	**Key Secondary Endpoint** • Improvement in fibrosis by at least 2 stages (NASH CRN system) AND no worsening of steatohepatitis (no worsening of lobular inflammation or hepatocellular ballooning grade 2005) on liver histology at Month 12 relative to the screening biopsy

Objectives	Endpoints
Secondary	**Secondary Endpoints**
• To evaluate the effect of CVC compared to placebo on liver histology at Month 12 relative to the Screening biopsy for the proportion of subjects with improvement in fibrosis by at least 1 stage (NASH CRN system), regardless of effect on steatohepatitis	• Improvement in fibrosis by at least 1 stage (NASH CRN system), regardless of effect on steatohepatitis, at Month 12 relative to the Screening biopsy
• To evaluate the effect of CVC compared to placebo on liver histology at Month 12 relative to the screening biopsy for the proportion of subjects with improvement in fibrosis by at least 2 stages (NASH CRN system), regardless of effect on steatohepatitis	• Improvement in fibrosis by at least 2 stages (NASH CRN system), regardless of effect on steatohepatitis, at Month 12 relative to the Screening biopsy
• To evaluate the safety and tolerability of CVC for the treatment of liver fibrosis in adult subjects with NASH	• Safety endpoints: incidence of adverse events, clinical adverse events, including major adverse cardiovascular events (MACE) and new-onset T2DM; laboratory data (hematology, chemistry, and fasting metabolic parameters), vital signs, and 12 lead ECG.
[redacted]	

[redacted]

Global Part 2:

Objectives: Global Part 2	Endpoints: Global Part 2
Primary • Demonstrate the superiority of CVC compared to placebo on the composite endpoint of histopathologic progression to cirrhosis, liver-related clinical outcomes, and all-cause mortality, as measured by the time to first occurrence of any of the listed adjudicated events (clinical outcomes composite endpoint) – (all subjects)	**Primary Endpoint** Time to first occurrence of any of the following adjudicated events: • Death (all cause) • Histopathologic progression to cirrhosis (i.e., NASH CRN fibrosis stage 4) • Liver transplant • MELD score ≥15 • Ascites (requiring intervention, i.e., large volume paracentesis ≥1L or initiation of a diuretic) • Hospitalization (as defined by a stay of ≥ 24 hours) for onset of: variceal bleed, hepatic encephalopathy (defined by a West Haven Stage of ≥2), spontaneous bacterial peritonitis (confirmed by diagnostic paracentesis with positive ascitic fluid bacterial culture) Each component of this endpoint will be considered by the independent adjudication committee, and only events confirmed by the committee will be included in the primary analysis. If any of the above events occurs in Part 1, they will be included in this primary efficacy endpoint for Part 2.
Secondary • To evaluate the effect of CVC compared to placebo on liver histology at Month 12 relative to the Screening biopsy for the proportion of subjects with improvement in fibrosis by at least 1 stage (NASH CRN system) AND no worsening of steatohepatitis (no worsening of lobular inflammation or hepatocellular ballooning grade 2005) – (subjects newly randomized in Part 2)	**Secondary Endpoints** • Improvement in fibrosis by at least 1 stage (NASH CRN system) AND no worsening of steatohepatitis (no worsening of lobular inflammation or hepatocellular ballooning grade 2005) on liver biopsy at Month 12 relative to the Screening biopsy – (subjects newly randomized in Part 2)

Objectives: Global Part 2	Endpoints: Global Part 2
• To evaluate the effect of CVC compared to placebo on liver histology at Month 12 relative to the Screening biopsy for the proportion of subjects with improvement in fibrosis by at least 1 stage (NASH CRN system), regardless of effect on steatohepatitis – (subjects newly randomized in Part 2)	• Improvement in fibrosis by at least 1 stage (NASH CRN system), regardless of effect on steatohepatitis, on liver biopsy at Month 12 relative to the Screening biopsy – (subjects newly randomized in Part 2)
• To evaluate the effect of CVC compared to placebo on liver histology at Month 12 relative to the screening biopsy for the proportion of subjects with improvement in fibrosis by at least 2 stages (NASH CRN system) AND no worsening of steatohepatitis (no worsening of lobular inflammation or hepatocellular ballooning grade 2005) – (subjects newly randomized in Part 2)	• Improvement in fibrosis by at least 2 stages (NASH CRN system) AND no worsening of steatohepatitis (no worsening of lobular inflammation or hepatocellular ballooning grade 2005) on liver biopsy at Month 12 relative to the screening biopsy – (subjects newly randomized in Part 2)
• To evaluate the effect of CVC compared to placebo on liver histology at Month 12 relative to the screening biopsy for the proportion of subjects with improvement in fibrosis by at least 2 stages (NASH CRN system) regardless of effect on steatohepatitis – (subjects newly randomized in Part 2)	• Improvement in fibrosis by at least 2 stages (NASH CRN system) regardless of effect on steatohepatitis on liver biopsy at Month 12 relative to the screening biopsy – (subjects newly randomized in Part 2)
• To evaluate the effect of CVC compared to placebo on liver histology at Month 60 relative to the Screening biopsy for the proportion of subjects with improvement in fibrosis by at least 1 stage (NASH CRN system) AND no worsening of steatohepatitis (no worsening of lobular inflammation or hepatocellular ballooning grade 2005) – (all subjects)	• Improvement in fibrosis by at least 1 stage (NASH CRN system) AND no worsening of steatohepatitis (no worsening of lobular inflammation or hepatocellular ballooning grade 2005) on liver biopsy at Month 60 relative to the Screening biopsy – (all subjects)

Objectives: Global Part 2	Endpoints: Global Part 2
• To evaluate the effect of CVC compared to placebo on liver histology at Month 60 relative to the Screening biopsy for the proportion of subjects with improvement in fibrosis by at least 1 stage (NASH CRN system), regardless of effect on steatohepatitis – (all subjects)	• Improvement in fibrosis by at least 1 stage (NASH CRN system), regardless of effect on steatohepatitis, on liver biopsy at Month 60 relative to the Screening biopsy – (all subjects)
• To evaluate the effect of CVC compared to placebo on liver histology at Month 60 relative to the screening biopsy for the proportion of subjects with improvement in fibrosis by at least 2 stages (NASH CRN system) AND no worsening of steatohepatitis (no worsening of lobular inflammation or hepatocellular ballooning grade 2005) – (all subjects)	• Improvement in fibrosis by at least 2 stages (NASH CRN system) AND no worsening of steatohepatitis (no worsening of lobular inflammation or hepatocellular ballooning grade 2005) on liver biopsy at Month 60 relative to the screening biopsy – (all subjects)
• To evaluate the effect of CVC compared to placebo on liver histology at Month 60 relative to the screening biopsy for the proportion of subjects with improvement in fibrosis by at least 2 stages (NASH CRN system), regardless of effect on steatohepatitis – (all subjects)	• Improvement in fibrosis by at least 2 stages (NASH CRN system), regardless of effect on steatohepatitis on liver biopsy at Month 60 relative to the screening biopsy – (all subjects)
• To evaluate the safety and tolerability of CVC for the treatment of liver fibrosis in adult subjects with NASH – (all subjects)	• Safety endpoints: incidence of adverse events, clinical adverse events, including major adverse cardiovascular events (MACE) and new-onset T2DM; laboratory data (hematology, chemistry, and fasting metabolic parameters), vital signs, and 12 lead ECG – (all subjects)

[redacted]

[redacted]

[redacted]

EU-specific changes that differ from the global objectives and endpoints affect the Part 1 objectives and endpoints. The EU objectives and endpoints are as follows:

Objectives: EU Part 1	Endpoints: EU Part 1
Primary • To demonstrate the superiority of CVC compared to placebo on liver histology at Month 12 relative to the Screening biopsy	**Primary Endpoint** • Improvement in fibrosis by at least 2 stages (NASH CRN system) AND no worsening of steatohepatitis (no worsening of lobular inflammation or hepatocellular ballooning grade 2005) on liver histology at Month 12 relative to Screening biopsy
Key Secondary None	**Key Secondary Endpoint** None
Secondary	**Secondary Endpoints**
[redacted]	

EU Part 2:

Objectives: EU Part 2	Endpoints: EU Part 2
Primary Same as global objectives	**Primary Endpoint** Same as global endpoint

[redacted]

4.3 Schedule of Assessments

Table 4-1 Schedule of Assessments

Part 1

[redacted]

[redacted]

[redacted]

[redacted]

Part 2

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

5. Statistical Methodology and Study Endpoints

5.1 Statistical Methods Planned in the Protocol and Determination of Sample Size

This SAP will be approved prior to database lock. The SAP expands the statistical section of the protocol and contains a detailed description of methods to analyze data collected in the study. The text portion of the SAP will be included in the CSR report as Appendix 16.1.9. Information on the PK, PK/PD, pharmacogenetic, and biomarkers analyses can be found in the PK data analysis plan. Information regarding health economics and outcomes research are found in a separate SAP.

5.1.1 Statistical and Analytical Plans

Statistical analyses will be conducted using SAS Version 9.4 or newer.

5.1.1.1 Common Conventions

5.1.1.1.1 Analysis Populations

The analysis populations will consist of subjects as defined below:

Table 5-1 Analysis Populations

Population	Definition	Study Treatment
Screened	All screened subjects who sign informed consent.	—
Intent-to-Treat (ITT)	All randomized subjects will be included in the randomly assigned treatment group, even if they receive a different treatment from that to which they are randomly assigned. This will be a supportive analysis set for selected efficacy analyses. Subjects with missing Screening biopsy data should not be randomized in the study, and hence will not be included in the ITT analysis set, should such subjects exist.	Randomized assignment
Modified Intent-to-Treat (mITT)	All randomized subjects who receive ≥ 1 dose of study drug will be included in the randomly assigned treatment group, even if they receive a different treatment from that to which they are randomly assigned. This will be the primary analysis set for the efficacy analyses.	Randomized assignment
Month 12 per protocol (PP1)	All subjects randomly assigned to a treatment group who have an evaluable biopsy at Screening and another after at least 6 months but before 15 months of follow-up, receive at least 6 months of assigned study drug and have no significant protocol deviations that potentially influence the primary efficacy assessment will be included in the PP1 analysis set. This will be a supportive analysis set for selected efficacy analyses of Part 1 and Part 2.	Randomized assignment

Population	Definition	Study Treatment
Month 60 per protocol (PP2)	All subjects randomly assigned to a treatment group who have an evaluable biopsy at Screening, have no significant protocol deviations that potentially influence the primary efficacy assessment, and either: • Have an adjudicated event that is a component of the primary composite endpoint within 24 months after randomization **OR** • Have at least 24 months of follow-up for clinical events **AND** one of the following: • Have an adjudicated event that is a component of the primary composite endpoint within 6 months of last intake of study drug **OR** • Have a biopsy at least 24 months after randomization and within 6 months of last intake of study drug. This will be a supportive analysis set for selected efficacy analyses of Part 2.	Randomized assignment
Safety	All subjects who received ≥ 1 dose of study drug will be included in the safety analysis set. Subjects will be included in the treatment group according to actual treatment received, even if they receive a different treatment than that to which they are randomly assigned. Subjects who inadvertently receive both treatments will be included with the CVC group. This will be the primary analysis set for the safety analyses.	Actual received[1]
Pharmacokinetic	All subjects who are randomized and receive at least one dose of study drug and have at least one post-dose PK sample will be included in the PK analysis set.	Randomized assignment

[1] Subjects will be summarized according to the study treatment received.

5.1.1.1.2 Study Treatments

The following treatment groups are defined for this study:

- Arm A: CVC 150 mg, once daily
- Arm B: Placebo once daily

5.1.1.1.3 Statistical Methodology

The methodologies defined below apply as specified to individual endpoints defined in this SAP. When indicated, 95% two-sided confidence intervals and two-sided p-values will be presented. Individual endpoint analyses may specify exceptions (additions or deletions) to these general definitions.

Table 5-2 Statistical Methodology

Methodology	Description

Categorical descriptive statistics	• Number and percentage of subjects in individual categories
	  ○ Subjects with ≥ 1 qualifying event counted once per individual category
	• N1 if denominator ≠ number of subjects in the population (standard percentage denominator)
	  ○ N1 = subjects with non-missing value at corresponding visit

Continuous descriptive statistics	• N1, mean, standard deviation (SD), median, minimum, maximum • N1 = subjects with non-missing value
CFB descriptive statistics	• Continuous descriptive statistics for baseline, post-baseline, and change from baseline (CFB) values • N1 = subjects with non-missing values at both baseline and the specified post-baseline analysis visit

CFB ANCOVA	• Continuous descriptive statistics and standard error (SE) for baseline, post-baseline, and CFB values
	• Estimates derived from mixed model for CFB value controlling for factors (treatment group, fibrosis stage [2 vs 3], presence or absence of T2DM) and covariate (baseline value)
	  ○ Least squares (LS) means and SEs
	  ○ LS mean differences, SEs, and confidence intervals (CIs) vs Placebo
	  ○ P-values from contrast t-test comparing treatment group vs Placebo
	• N1 = subjects with non-missing values at both baseline and the specified post-baseline analysis visit

Responder	• Categorical descriptive statistics for responders and nonresponders
	  ○ Nonresponders include:
	  ▪ Subjects who do not meet responder criteria
	  ▪ Subjects missing the post-baseline liver biopsy will be included as nonresponders (see Section 5.1.1.1.5 for handling of missing biopsy or inadequate biopsy)
	• Risk differences and Wilson's method for CIs vs Placebo
	• Estimates derived from Cochran-Mantel-Haenszel (CMH) model controlling for factors (treatment group, fibrosis stage at baseline [2 vs 3] and presence or absence of T2DM at baseline)
	  ○ Mantel-Haenszel (MH) odds ratios (ORs) and 95% CIs vs Placebo
	  ○ P-values comparing treatment group vs Placebo
	• N1 = all subjects unless otherwise specified

Time-to-event (TTE)	• Categorical descriptive statistics for subjects with events and censoring
	  ○ Censoring includes:
	  ▪ Subjects who do not meet event criteria
	  ▪ Subjects with no post-baseline values of corresponding parameters
	• Quartiles, estimated event rates, and CIs derived from Kaplan-Meier (KM) nonparametric model using log-log transformation of survival function using product-limit method
	• Estimates derived from Cox proportional hazards model controlling for factors (treatment group, fibrosis stage at baseline [2 vs 3] and presence or absence of T2DM at baseline)
	  ○ Hazard ratios (HRs) and CIs vs Placebo
	• Estimates derived from stratified log-rank model controlling for factors (treatment group, fibrosis stage at baseline [2 vs 3] and presence or absence of T2DM at baseline)
	  ○ P-values comparing treatment group vs Placebo
	• N1 = all subjects unless otherwise specified

Methodology	Description
KM figure	• Step-function figure of cumulative distribution function [1 - survival function] estimates with censoring indicators, derived from KM nonparametric model

CFB = change from baseline; ANCOVA = analysis of covariance; TTE = time-to-event.

Raw and derived data listings will be provided, and will be fully defined in the table, figure, and data listing specification document.

5.1.1.1.4 Estimands

Estimands for Global Part 1 primary, key secondary, and secondary endpoints are provided below:

Table 5-3 Estimand Language

Endpoint: Global Part 1	Estimand Language
Primary	• **Population**: All randomized subjects with NASH CRN fibrosis stage 2 or 3 who also met inclusion and exclusion criteria for the study. The analysis population is the modified ITT population. • **Variable/Endpoint**: Improvement in fibrosis by at least 1 stage (NASH CRN system) AND no worsening of steatohepatitis (no worsening of lobular inflammation or hepatocellular ballooning grade 2005) on liver histology at Month 12 relative to Screening biopsy • **Intercurrent Event**: Hypothetically envisage that those who withdraw from the study prior to the liver biopsy at Month 12 or those who do not have readable liver biopsies at the Month 12 visit assessment will be considered as non-responders (those who do not improve in fibrosis stage). • **Population-level Summary**: Difference in responder rates from screening biopsy for fibrosis stage improvement (with no worsening of steatohepatitis) at Month 12 between the CVC group versus the placebo group using the Cochran-Mantel-Haenszel (CMH) test, stratified by NASH CRN fibrosis stage (2 vs 3) and T2DM at baseline (presence vs absence)
Key Secondary	• **Population**: All randomized subjects with NASH CRN fibrosis stage 2 or 3 who also met inclusion and exclusion criteria for the study. The analysis population is the modified ITT population. • **Variable/Endpoint**: Improvement in fibrosis by at least 2 stages (NASH CRN system) AND no worsening of steatohepatitis (no worsening of lobular inflammation or hepatocellular ballooning grade 2005) on liver histology at Month 12 relative to Screening biopsy • **Intercurrent Event**: Hypothetically envisage that those who withdraw from the study prior to the liver biopsy at Month 12 or those who do not have readable liver biopsies at the Month 12 visit assessment will be considered as non-responders (those who do not improve in fibrosis stage). • **Population-level Summary**: Difference in responder rates from screening biopsy for fibrosis stage improvement (with no worsening of steatohepatitis) at Month 12 between the CVC group versus the placebo group using the Cochran-Mantel-Haenszel (CMH) test, stratified by NASH CRN fibrosis stage (2 vs 3) and T2DM at baseline (presence vs absence)

Endpoint: Global Part 1	Estimand Language
Secondary	• **Population**: All randomized subjects with NASH CRN fibrosis stage 2 or 3 who also met inclusion and exclusion criteria for the study. The analysis population is the modified ITT population. • **Variable/Endpoint**: Improvement in fibrosis by at least 1 stage (NASH CRN system), regardless of effect on steatohepatitis, at Month 12 relative to Screening biopsy • **Intercurrent Event**: Hypothetically envisage that those who withdraw from the study prior to the liver biopsy at Month 12 or those who do not have readable liver biopsies at the Month 12 visit assessment will be considered as non-responders (those who do not improve in fibrosis stage). • **Population-level Summary**: Difference in responder rates from screening biopsy for fibrosis stage improvement at Month 12 between the CVC group versus the placebo group using the Cochran-Mantel-Haenszel (CMH) test, stratified by NASH CRN fibrosis stage (2 vs 3) and T2DM at baseline (presence vs absence) The other secondary endpoint (same as above secondary endpoint but with improvement by "at least 2 stages") will be analyzed similarly.

The EU-specific estimands will be as follows:

- EU primary estimand: same as the Global Part 1 key secondary estimand
- EU key secondary estimand: none
- EU secondary estimands: same as the Global Part 1 secondary estimands

5.1.1.1.5 Missing Data

General missing data handling conventions are specified for methodologies in Section 5.1.1.1.3 and summarized as follows:

Table 5-4 Missing Data Handling by Endpoint Type

Parameter type	Timing	Missing Data Handling
Responder (Part 1: primary endpoint at Month 12)	Screening, Month 12	• All subjects in mITT analysis set included. Subjects with missing baseline biopsy data should not be randomized in the study. • Any available adequate liver biopsy after baseline will be used as the biopsy for the target visit (ie., Month 12), no matter when it was obtained relative to the first dose of study drug. • Subjects with multiple liver biopsies will have the evaluable biopsy closest to the target visit included in the analysis. • Subjects who do not have an evaluable liver biopsy at both Screening and the target visit will be included in the analyses as nonresponders. • Summaries of subjects who do and who do not have evaluable liver biopsies at the target visit will be provided to assess whether demographics, baseline data or safety data are predictive of missingness.

Parameter type	Timing	Missing Data Handling
		• Subjects who do not have an evaluable biopsy at the target visit will be those with no post baseline liver biopsy or those with a biopsy reading of non-evaluable from the central reader. • Sensitivity analyses for the primary efficacy endpoint will be performed, accounting for both missing at random and missing not at random scenarios.
Responder (Part 2: secondary endpoint at Month 60)	Screening, Month 60	• Same as above for target visit, Month 60
Time-to-event (Part 2: primary endpoint)	Treatment Period	• All subjects in mITT analysis set included. • Any available adequate liver biopsy after baseline will be used for assessment of components of the clinical endpoint that require a biopsy, no matter when obtained. • Subjects with multiple liver biopsies will be included as an event if any (one or more) biopsy meets the definition of the primary endpoint, as determined by the adjudication committee. • Subjects who do not have an evaluable liver biopsy at both Screening and post-baseline will be included in the analyses as having an event only if they meet a component that does not require a biopsy; otherwise they will be included as not having an event and censored at the last visit. • Subjects with partial follow-up data will be included as an event if they meet the criteria for any component of the primary efficacy endpoint, and as a nonevent if they do not meet the criteria for any component. • A sensitivity analysis assuming data are missing not at random will be reported. • An additional sensitivity analysis imputing an event at the time of censoring will also be reported, although it is anticipated that this analysis will not be informative unless many subjects discontinue due to unobserved NASH progression, which cannot be known.
CFB ANCOVA	Months 6 and 12 in Part 1, and Month 12 and annually thereafter in Part 2 (Month 6 in Part 2 for some endpoints)	• If missing covariates (including baseline if applicable) ○ Subject excluded • Observed data for mean absolute change and relative change from baseline to the times specified in the endpoints will be analyzed. There is no imputation for missing data.
Multiple imputation (Part 1: primary efficacy endpoint)	Month 12 MAR and MNAR	A sensitivity analysis for the primary efficacy analysis in Part 1 will use multiple imputation to investigate the impact of missing data under a Missing At Random (MAR) assumption. For this analysis, the Month 12 biopsy results will be imputed for the subjects who do not have an evaluable biopsy after baseline. The imputation model will be defined before Part 1 unblinding to avoid the addition bias. The multiple imputation model will use the MI procedure in SAS, using code such as the following: proc mi data=<inputdsn> seed=1305417 nimpute=100 out=<outputdsn>; class outcome trt; var trt <explanatory> outcome;

Parameter type	Timing	Missing Data Handling

		monotone logistic (outcome=trt <explanatory>);
		run;
		
		The variables in the imputation model ("explanatory" in the pseudocode above) will be defined before the database is locked for Part 1 and documented in an amendment to the SAP. Explanatory variables are expected to include baseline values of age, weight change from baseline to Month 12, BMI, sex, race, site, and other baseline values, stratification factors of fibrosis stage at baseline [2 vs 3] and T2DM at baseline [presence vs. absence], and post-randomization variables that correlate with the Part 1 primary endpoint among subjects with complete data based on review before unblinding.
		
		Data from each imputation will be analyzed by CMH using code such as the following:
		
		PROC FREQ DATA=<outputdsn>;
		TABLES stratification * trt * outcome / CMH;
		ODS OUTPUT CMH=cmh;
		BY _Imputation_;
		RUN;
		
		Data from the multiple imputations will be combined with Rubin's rule after application of the Wilson-Hilferty transformation. Each imputation will be analyzed using the primary analysis method (CMH) and the results combined with the MIANALYZE procedure in SAS. The p-value for treatment will be reported.
		
		SAS code such as the following will accomplish this analysis:
		** Apply W-H transformation;
		DATA cmh_wh;
		SET cmh(WHERE=(AltHypothesis="General Association"));
		cmh_value_wh=((VALUE/DF)**(1/3) - (1-2/(9*DF)))/ SQRT(2/(9*DF));
		cmh_sterr_wh = 1.0;
		RUN;
		
		*** Combine results;
		PROC MIANALYZE DATA=cmh_wh;
		ODS OUTPUT PARAMETERESTIMATES=mian_cmh_wh;
		MODELEFFECTS cmh_value_wh;
		STDERR cmh_sterr_wh;
		RUN;
		
		*** Compute one-sided p-value;
		DATA mian_cmh_wh_p; SET mian_cmh_wh;
		IF tValue > 0 THEN Probt_upper = Probt/2;
		ELSE Probt_upper = 1-Probt/2;
		RUN;
		
		A sensitivity analysis for the primary efficacy analysis in Part 1 will use pattern-mixture model (PMM) with a control-based pattern imputation (or jump to reference) to investigate the impact of missing data under a Missing Not At Random (MNAR) assumption. First, we will impute missing data at time-point Month 12. In order to do it using

Parameter type	Timing	Missing Data Handling

		the control-based imputation method, we will separate our data into two datasets: DATAIN_MONO_IMP12, containing all placebo subjects and those subjects from the CVC arm that have values at time-point Month 12 missing; and DATAIN_MONO_REST12, containing the rest of the subjects from the CVC arm (those with non-missing SCORE_12).
		
		We will now call PROC MI to impute missing data at time-point Month 12 based on the model estimated exclusively from placebo subjects with non-missing values at time-point Month 12.
		
		proc mi data=DATAIN_MONO_IMP12 seed=234 nimpute=100 out= DATAIN_REG_IMP12
		class outcome ;
		var <explanatory> outcome;
		monotone logistic (outcome= <explanatory>);
		run;
		
		The following data step assembles back a dataset containing all subjects.
		data DATAIN_IMP1;
		set DATAIN_MONO_REST1 DATAIN_REG_IMP1;
		run;
		
		Data from each imputation will be analyzed using CMH and results will be combined using following codes.
		
		PROC FREQ DATA=DATAIN_IMP1;
		TABLES stratification * trt * outcome / CMH;
		ODS OUTPUT CMH=cmh2;
		BY _Imputation_;
		RUN;
		
		DATA cmh_wh2;
		SET cmh2(WHERE=(AltHypothesis="General Association"));
		cmh_value_wh=((VALUE/DF)**(1/3) - (1-2/(9*DF)))/ SQRT(2/ (9*DF));
		cmh_sterr_wh = 1.0;
		RUN;
		
		*** Combine results;
		PROC MIANALYZE DATA=cmh_wh2;
		ODS OUTPUT PARAMETERESTIMATES=mian_cmh_wh2;
		MODELEFFECTS cmh_value_wh;
		STDERR cmh_sterr_wh;
		RUN;
		
		*** Compute one-sided p-value;
		DATA mian_cmh_wh_p; SET mian_cmh_wh2;
		IF tValue > 0 THEN Probt_upper = Probt/2;
		ELSE Probt_upper = 1-Probt/2;
		RUN;

Parameter type	Timing	Missing Data Handling

Multiple imputation (Part 2: primary efficacy endpoint)	Treatment Period	Informative Censoring Imputation Model:[6]
		
		Let k index the censoring times before t*. (Note: t*≈ 96 months for subjects randomized in Part 1 and 84 months newly randomized subjects in Part 2.)
		Let $t_{k,0}$ denote the latest failure time prior to c_k (or equal to it) when $t_1 \leq c_k$.
		Let $t_{k,j}$ denote the jth failure time after c_k, j = 1, 2, … , J_k, when $c_k < t_M$. Note that the possible values of J_k range from 1 to M, depending on the position of c_k with respect to the order of the t_m's (m = 1, 2, … , M): J_k equals M if $c_k < t_1$, and J_k equals 1 if $t_{M-1} = c_k < t_M$. From the observed data, we obtain the Kaplan–Meier estimates $\hat{S}(t)$ for the survival distribution for the event times, and it has support on the observed event times (t_1, t_2, … , t_M).
		
		The estimate for the conditional cumulative incidence function for a subject with early discontinuation at c_k to have the event by the time t in $[t_{k,j} < t < t_{k,j+1}]$, for j = 1, 2, 3, … , J_k with $t_{k,J_k+1} = t^*$ can be obtained as follows.
		
		$$\hat{F}_{j,k}(\theta) = 1 - \frac{\hat{S}(t_{k,j+1})^{\theta}}{\hat{S}(c_k)^{\theta}}$$
		
		The estimate for the conditional probability for a subject with early discontinuation at c_k to have the event by the time t in $[c_k < t < t^*]$, can be obtained as follows.
		
		$$\hat{f}_{j,k}(\theta) = \frac{\hat{S}(c_k)^{\theta} - \hat{S}(t*)^{\theta}}{\hat{S}(c_k)^{\theta}}$$
		
		Under this formulation, θ > 1 (or < 1) implies a higher (or lower) hazard after c_k for patients with premature discontinuation at c_k than for patients with continued follow-up after c_k. The test treatment group would usually have $\theta_{active} > \theta_{placebo}$ specified, and with $\theta_{Placebo} = 1$, $\theta = (\theta_{active} / \theta_{Placebo}) = \theta_{active}$ becomes a single parameter for calibrating sensitivity analyses. The value of θ_{active} will be finalized before database lock.
		
		*** To obtain the survival function estimates $\hat{S}(t_{k,j+1})$ and $\hat{S}(c_k)$ to generate $\hat{f}_{j,k}(\theta)$ and $\hat{F}_{j,k}(\theta)$:
		
		proc lifetest data=xxx OUTSURV=SURVEST;
		time AVAL*CNSR(1);
		strata ARMCD / test=logrank;
		SURVIVAL CONFTYPE=LINEAR;
		run;
		
		In SURVEST datasets, SURVIVAL column is what we need to obtain the survival function estimates. Please note that $\hat{F}_{j,k}(\theta)$ is from 0 to 1. The multiple imputation scheme is as follows:
		1. Generate a random number p from the uniform distribution between 0 and 1, and for computational convenience, use linear interpolation to impute failure times.

Parameter type	Timing	Missing Data Handling
		Use UNIFORM function in SAS to generate a pseudorandom number p from the uniform distribution on [0, 1] for the following scenario. 2. Suppose a subject has early discontinuation at time c_k, perform the following failure time imputation: a. If $0 \le p \le \hat{F}_{k,0}(\theta)$, then impute the failure time $t_k^{(l)}$ between c_k and $t_{k,1}$ as $c_k + (t_{k,1} - c_k)\frac{p}{\hat{F}_{k,0}(\theta)}$, where l indicates the lth imputation set. b. If $\hat{F}_{k,j}(\theta) \le p \le \hat{F}_{k,j+1}(\theta)$, for j = 0, 1, 2, 3, … , (J_k - 1), then impute failure time $t_k^{(l)}$ between $t_{k,j+1}$ and $t_{k,j+2}$ as $\left(t_{k,j+1} + (t_{k,j+2} - t_{k,j+1}) * \frac{p - \hat{F}_{k,j}(\theta)}{\hat{F}_{k,j+1}(\theta) - \hat{F}_{k,j}(\theta)}\right)$, where $t_{k,J_k+1} = t^*$. c. If $p > \hat{F}_{k,J_k}(\theta)$, then manage the patient as having no event by the end of follow-up time t*. 3. Suppose a patient has premature discontinuation between t_M and t*, so that ($t_M < c_k < t^*$): a. If $p \le \hat{f}_{j,k}(\theta)$, then impute failure time $t_k^{(l)}$ between c_k and t* as $c_k + (t^* - c_k)\frac{p}{\hat{f}_{k,0}(\theta)}$; b. Otherwise, manage the patient as having no event by the end of follow-up time t*. The imputation is performed separately for the patients with premature discontinuation in each treatment group. The above imputation procedure is repeated to form L imputed data sets. *** <u>Combining Chi-square distributed statistics after multiple imputation:</u> After multiple imputation, it is possible to combine the test results from all available tests in PROC LIFETEST such as log-rank test. Each of these tests is based on a chi-square distributed statistic that can be normalized using a Wilson-Hilferty transformation. If X^2 is chi-square distributed with d degrees of freedom, then $$W = \sqrt[3]{\frac{X^2}{d}} \approx N\left(1 - \frac{2}{9d}, \frac{2}{9d}\right), or$$ $$\frac{\sqrt[3]{X^2/d} - (1 - \frac{2}{9d})}{\sqrt{\frac{2}{9d}}} \approx N(0,1)$$ Test statistics and their corresponding degrees of freedom can be found in the ODS output dataset HomTests produced by PROC LIFETEST when using the STRATA statement with TEST option. These test

Parameter type	Timing	Missing Data Handling

		statistics obtained from each imputed dataset can be normalized using the above formula and the normalized values can let be input for PROC MIANALYZE in order to obtain combined results for these chi-square based tests.
		
		DATA _tmc_tests1;
		set _tmc_tests;
		*** ChiSq is a chi-square distributed statistic (e.g., logrank) and DF is its corresponding degrees of freedom as produced by PROC LIFETEST ***;
		ChiSq_wh=((ChiSq/DF)**(1/3) - (1-2/(9*DF)))/SQRT(2/(9*DF));
		ChiSq_stderr_wh = 1.0;
		run;
		
		proc sort data=_tmc_tests1;
		BY _Imputation_;
		run;
		
		proc MIANALYZE data=_tmc_tests1;
		modeleffects ChiSq_wh;
		stderr ChiSq_stderr_wh;
		ods output parameterestimates=resprefix._tests;
		run;
		
		*** Combining results of Cox regression and exponentiation of the combined log-hazard ratios:
		Cox regression model parameter estimates from PROC PHREG represent log hazard ratios. They are normally distributed and do not need any further transformation to be combined with Rubin's rules using PROC MIANALYZE. However, the end results need to be exponentiated to obtain the numerical values of the hazard ratios and associated confidence intervals.
		
		proc phreg data=_tmc_1;
		by _Imputation_;
		class trtvar <classvars>;
		model timevar * censvar(censval) = trtvar covars / risklimits ties=efron rl;
		ods output ParameterEstimates=_es;
		run;
		
		/* Combine model coefficients. */
		proc sort data=_es;
		by Parameter ClassVal0 _Imputation_;
		run;
		proc MIANALYZE data=_es;
		by Parameter ClassVal0;
		modeleffects Estimate;
		stderr stderr;
		ods output ParameterEstimates = _es_mianal;
		run;
		
		/*Exponentiate in order to obtain hazard ratio estimates and confidence intervals */
		data resprefix._hr;

Parameter type	Timing	Missing Data Handling
		set _es_mianal; Log_HR_comb=Estimate; HR_comb=exp(Estimate); HR_LCL_comb=exp(LCLMean); HR_UCL_comb=exp(UCLMean); keep Parameter ClassVal0 Log_HR_comb HR_comb HR_LCL_comb HR_UCL_comb Probt; rename Probt=HR_pval_comb; run;

5.1.1.1.6 Site Pooling

The following section applies when statistical analyses are to be performed with site as a stratification factor (e.g., multiple imputation model for Part 1).

All small sites within a country with fewer than 5 ITT subjects in a treatment group will be pooled from largest to smallest until the pooled site has ≥5 ITT subjects in each treatment group. After the pooling, any remaining unpooled sites will be pooled with the smallest existing site.

If any sites need to be pooled, then any analysis performed by trial site will be performed by pooled site instead.

5.1.1.1.7 Other Common Conventions

Continuous data will be described with sample size, mean, standard deviation, median, minimum and maximum. Categorical data will be described with frequencies and percentages in each category. Two-sided p-values for all hypothesis tests will be reported and will be interpreted in accordance with the multiple comparisons procedure described in Section 5.1.1.3.3.

Subjects will be enrolled in 2 parts, Part 1 and Part 2. Only subjects randomized for Part 1 will contribute to the Part 1 analyses, and all randomized subjects (i.e., subjects randomized for Part 1 and subjects randomized for Part 2) will contribute to the Part 2 analyses. When subjects who contribute to Part 1 have completed their 12-month biopsy (or have been followed for a sufficient time to confirm that such a biopsy will not be obtained), the Part 1 analysis will commence. The Part 1 analysis will summarize subject disposition, significant protocol deviations, study intervention exposure, concomitant medications, demographics and baseline characteristics, efficacy through the Month 12 visit, and safety for subjects enrolled in Part 1, including all available safety data in the database at the time of the Part 1 database lock. At the completion of the entire study, analyses will be provided for all subjects randomized, including all available data in the database.

5.1.1.2 Demographics

5.1.1.2.1 Analysis Populations

The distribution of subjects within the analysis populations will be summarized as follows:

Table 5-5 Analysis Population Summaries

Population	Description	Analyses	Methodology
Screened population	Distribution overall	Part 1 and Part 2 analyses	Categorical counts
ITT, mITT, and Safety populations	Distribution in total and by treatment group	Part 1 and Part 2 analyses	Categorical counts
Month 12 Per Protocol (PP1) population	Distribution in total and by treatment group	Part 1 and Part 2 analyses	Categorical counts
Month 60 Per Protocol (PP2) population	Distribution in total and by treatment group	Part 2 analyses	Categorical counts
Pharmacokinetic population	Distribution in total and by treatment group	Part 1 and Part 2 analyses	Categorical counts

5.1.1.2.2 Participant Disposition

Subjects disposition encompasses the distribution of subjects who were screened, randomized, started study drug, completed the study, and discontinue study drug early (during Part 1 or Part 2), and discontinued from the study early (during Part 1 or Part 2), Subjects with Month 12 biopsy (evaluable, nonevaluable, and missing) for Parts 1 and 2 and with Month 60 biopsy (evaluable, nonevaluable, and missing) for Part 2 will also be summarized. Subject disposition will be summarized as follows:

Table 5-6 Participant Disposition Summaries

Parameter	Description	Timing	Methodology
Screening disposition	Distribution in the Screened Population in total	Screening Period	Categorical descriptive statistics
Study disposition[1]	Distribution in the ITT Population and mITT Population in total and by treatment group and by region	Treatment Period	Categorical descriptive statistics

[1] Subjects who prematurely discontinued study drug and/or study will be listed.

5.1.1.2.3 Protocol Deviations

Significant protocol deviations will be summarized for the ITT population (i.e., all randomized subjects) as follows:

Table 5-7 Protocol Deviation Summary

Parameter	Description	Timing	Methodology
Significant protocol deviations	Distribution in the ITT Population in total and by treatment group The set of subjects excluded from the PP1 analysis set will be identified prior to database lock and unblinding the study database for Part 1. Additional subjects may be included or excluded in the PP1 set for analysis of Part 2 if subjects are randomized to Part 2 after the analysis of Part 1 or if additional significant protocol deviations are identified between Part 1 analysis and unblinding of the study database for Part 2. The set of subjects excluded from the PP2 analysis set will be identified prior to database lock and unblinding the study database for Part 2.	—	Categorical descriptive statistics

5.1.1.2.4 Demographics

Demographics will be summarized in total and by treatment group for the ITT and mITT populations, as follows:

Table 5-8 Demographic Summaries

Parameter	Description	Timing	Methodology
Age	Age (years) relative to informed consent date	Informed consent	Continuous descriptive statistics
Age group	• < 65 years • ≥ 65 years	Informed consent	Categorical descriptive statistics

Sex, race, and ethnicity	• Sex
	  ○ Male
	  ○ Female
	• Race group
	  ○ White
	  ○ Non-white
	  ▪ Black or African American
	  ▪ Asian
	  ▪ American Indian or Alaska Native
	  ▪ Native Hawaiian or Other Pacific Islander
	• Ethnicity
	  ○ Hispanic or Latino
	  ○ Not Hispanic or Latino
		Screening Period	Categorical descriptive statistics

5.1.1.2.5 Baseline Characteristics

Baseline characteristics will be summarized in total and by treatment group for the ITT and mITT populations as follows:

Table 5-9 Baseline Characteristics Summaries

Parameter	Description	Timing	Methodology
Baseline Demographic characteristics	• Height (m) • Weight (kg) • Body mass index (BMI) ○ Weight (kg) / height $(m)^2$ • Waist circumference • Hip circumference • Waist-to-hip ratio	Baseline visit, except for height and weight at Screening for determination of BMI	Continuous descriptive statistics

[redacted]

[redacted]

5.1.1.2.6 Medical History

Medical history, encompassing abnormalities and surgeries reported as occurring before the Screening Visit, will be coded using the Medical Dictionary for Regulatory Activities (MedDRA), version 22.1 or newer. Unique subjects who are reported to have medical history events will be summarized by system organ class (SOC) and preferred term (PT) in total and by treatment group for the ITT Population as follows:

Table 5-10 Medical History Summary

Parameter	Description	Timing	Methodology
Medical history	Abnormalities and surgeries occurring before the Screening Visit	Screening Period	Categorical descriptive statistics

SOCs will be sorted alphabetically; PTs will be sorted in descending frequency.

5.1.1.2.7 Prior and Concomitant Medications

Medications will be coded using the World Health Organization (WHO) Drug Dictionary, version MAR2018 or newer. Unique subjects who reported medications will be summarized by Anatomical Therapeutic Chemical (ATC) (4^{th} level) drug class, drug code, and preferred drug name in total and by treatment group for the ITT Population, separately for prior and concomitant medications, as defined below:

Table 5-11 Medication Summaries

Parameter	Description	Timing	Methodology
Prior medications	Non-study medications taken ≥ 1 time 30 days before the study treatment start date and stopped before the study treatment start date.	Pretreatment Period	Categorical descriptive statistics
Concomitant medications	Non-study medications taken after the study treatment start date through the 30-Day Follow-up visit after last study drug intake (or final visit, for subjects who do not complete a follow-up visit), regardless of medication start date • This includes medication ongoing at the time of first study drug intake and medications started after first study drug intake.	Treatment Period through 30-Day Follow-up visit after last study drug intake (or final visit, for subjects who do not complete a follow-up visit)	Categorical descriptive statistics

ATC4 classes will be sorted alphabetically; PTs will be sorted in descending frequency.

5.1.1.3 Efficacy Analyses

Efficacy analyses will be based on the mITT Population (primary analysis set). Selected efficacy analyses will be based on the ITT Population (supportive analysis set). Sensitivity analyses for the primary efficacy endpoint will be performed using the ITT and PP1 populations.

The following efficacy assessments and terms are defined:

Table 5-12 Efficacy Assessments

Assessment/Term	Description
Improvement of fibrosis stage (NASH CRN system) and no worsening of steatohepatitis (no worsening of lobular inflammation or hepatocellular ballooning grade 2005) on liver histology at Month 12 relative to the Screening biopsy	Liver biopsies will be performed per standard of care (e.g., taking into account the subject's health status, coagulation profile, and use of concomitant medications). Subjects who discontinue study drug are encouraged to continue with all other evaluations as scheduled in this protocol, including biopsies and other invasive procedures at the scheduled visits. If a subject discontinues study participation early and has received study drug for at least 6 months in Part 1 or Part 2, then a biopsy should be taken within 30 days of discontinuation, if feasible. All liver biopsies will be evaluated by an independent central pathologist; whenever possible, the same pathologist will evaluate all biopsies from an individual subject. The central pathologist will be blinded to individual treatment assignment. Liver biopsy at Screening and Months 12 and 60 will be used to determine improvement in histologic fibrosis stage (NASH CRN and modified Ishak systems). Evaluation of fibrosis stage will be based on the following NASH CRN Fibrosis Staging System. For analysis purposes, NASH CRN fibrosis stages 1, 1A, 1B, and 1C will all be considered as stage 1.

<table>
<tr><th>Assessment/Term</th><th>Description</th></tr>
<tr><td></td><td>
NASH CRN Fibrosis Staging System
<table>
<tr><th>Fibrosis</th><th>Stage</th></tr>
<tr><td>None</td><td>0</td></tr>
<tr><td>Perisinusoidal or periportal</td><td>1</td></tr>
<tr><td>Mild, zone 3, perisinusoidal</td><td>1A</td></tr>
<tr><td>Moderate, zone 3, perisinusoidal</td><td>1B</td></tr>
<tr><td>Portal/periportal</td><td>1C</td></tr>
<tr><td>Perisinusoidal and portal/periportal</td><td>2</td></tr>
<tr><td>Bridging fibrosis</td><td>3</td></tr>
<tr><td>Cirrhosis</td><td>4</td></tr>
</table>
NASH CRN Staging System
The NASH CRN adopted the staging system validated by Kleiner. The authors evaluated a scoring system comprising 14 histological features that addressed the full spectrum of lesions of NAFLD. Five features were evaluated semi-quantitatively (steatosis [0-3], location [0-3], lobular inflammation [0-3], hepatocellular ballooning [0-2], and fibrosis stage [0-4]) and 9 additional features were recorded as present or absent. As a result of this evaluation, the NASH CRN staging system for evaluating histological changes after therapeutic intervention trials adopted a defined and validated semiquantitative scoring system, NAS, based on the unweighted sum of steatosis, lobular inflammation, and hepatocellular ballooning scores (2005). The system is simple and requires only routine histochemical stains (hematoxylin and eosin and Masson trichrome stains). This scoring system and NAS for NAFLD and NASH were shown to have reasonable inter-rater reproducibility in both adults and children with any degree of NAFLD. A NAS of > 5 correlated with a diagnosis of NASH, and biopsies with scores of < 3 were diagnosed as not NASH. The primary purpose of the NAS is to assess overall histological change; it is not intended that numeric values replace the pathologist’s diagnostic determination of steatohepatitis.

The evaluation of steatohepatitis includes the following:
<table>
<tr><th>Item</th><th>Score</th><th>Extent</th></tr>
<tr><td rowspan="4">Lobular inflammation</td><td>0</td><td>No foci</td></tr>
<tr><td>1</td><td>< 2 foci/200x</td></tr>
<tr><td>2</td><td>2-4 foci/200x</td></tr>
<tr><td>3</td><td>> 4 foci/200x</td></tr>
<tr><td rowspan="3">Hepatocellular ballooning (2005 score)</td><td>0</td><td>None</td></tr>
<tr><td>1</td><td>Few ballooned cells</td></tr>
<tr><td>2</td><td>Many cells/prominent ballooning</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><th>Assessment/Term</th><th>Description</th></tr>
<tr><td>Modified Ishak fibrosis staging</td><td>As the original Ishak staging system was designed for chronic viral hepatitis[1], a modified Ishak system will be used to classify NASH fibrosis stage, as it provides additional granularity at higher fibrosis stages. Stages F0–F2 will use the same histologic definitions as the NASH CRN system (without further classification of Stage F1 into F1a, F1b, or F1c), and Stages F3–F6 will be defined as follows: occasional bridging fibrosis (< 50% linkage of portal and/or central zones) [Stage F3]; marked bridging fibrosis (> 50% linkage of portal and/or central zones but not yet cirrhosis) [Stage F4]; early or incomplete cirrhosis (Stage F5); and established or advanced cirrhosis (Stage F6)[2].
<table>
<tr><th>Architectural changes, Fibrosis and Cirrhosis</th><th>Score</th></tr>
<tr><td>No fibrosis</td><td>0</td></tr>
<tr><td>Perisinusoidal or periportal</td><td>1</td></tr>
<tr><td>Perisinusoidal and portal/periportal</td><td>2</td></tr>
<tr><td>Occasional bridging fibrosis (< 50% linkage of portal and/or central zones)</td><td>3</td></tr>
<tr><td>Marked bridging fibrosis (> 50% linkage of portal and/or central zones but not yet cirrhosis)</td><td>4</td></tr>
<tr><td>Early or incomplete cirrhosis</td><td>5</td></tr>
<tr><td>Established or advanced cirrhosis</td><td>6</td></tr>
</table>
</td></tr>
</table>

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

Baseline assessments for applicable efficacy endpoints defined as follows:

Table 5-13 Efficacy Endpoint Baseline Definitions

Type of Endpoint	Description	Timing
Part 1: Primary, secondary [redacted] Improvement of fibrosis stage (NASH CRN system) on liver histology at Month 12 relative to the Screening biopsy	Fibrosis stage (NASH CRN system) based on Screening biopsy	Screening biopsy result before Baseline visit and before subject is randomized
Part 1: Primary, secondary [redacted] Worsening of steatohepatitis (no worsening of lobular inflammation or hepatocellular ballooning grade 2005) on liver histology at Month 12 relative to the Screening biopsy	Steatohepatitis status (lobular inflammation or hepatocellular ballooning grade 2005) based on Screening biopsy	Screening biopsy result before Baseline visit and before subject is randomized
Part 1 and Part 2: Change in baseline in noninvasive assessments of liver fibrosis (FIB-4, APRI, NFS, and ELF score)	Occurs within 3 months before the Baseline visit. Baseline visit is defined as the last available assessment before the date of 1st dosing (for all efficacy and safety endpoints).	Screening serum hepatic fibrosis indices
Part 1 and Part 2: Change in baseline in noninvasive assessments of liver fibrosis (liver stiffness)	The Screening value (if not performed by TE within 6 months prior to the first day of Screening) will be used as the Baseline value. TE will be performed at Baseline if not performed within 6 months prior to the first day of Screening.	Screening via TE or liver imaging
Part 2: Primary efficacy endpoint: Time to first occurrence of adjudicated events	Date of Baseline visit will be used in the calculation for time to first occurrence of any of the following adjudicated event endpoints: • Death (all cause) • Histopathologic progression to cirrhosis (based on NASH CRN fibrosis stage 4) • Liver transplant • MELD score ≥15 • Ascites (requiring intervention, i.e., large volume paracentesis ≥1L or initiation of a diuretic) • Hospitalization (as defined by a stay of ≥ 24 hours) for onset of: variceal bleed, hepatic encephalopathy (defined by a West Haven Stage of ≥2), spontaneous bacterial peritonitis (confirmed by diagnostic paracentesis with positive ascitic fluid bacterial culture) If any of the adjudicated events occur in Part 1, they will be included in the primary efficacy endpoint for Part 2.	For subjects newly randomized in Part 2: Date of Baseline visit in Part 2 For subjects continuing from Part 1: Date of Baseline visit in Part 1

Type of Endpoint	Description	Timing
Part 2: Secondary efficacy endpoint (for subjects newly randomized in Part 2): Improvement of fibrosis stage (NASH CRN system) on liver histology at Month 12 relative to the Screening biopsy	Fibrosis stage (NASH CRN system) based on Screening biopsy.	Screening biopsy result before Baseline visit and before subject is randomized in Part 2
Part 2: Secondary efficacy endpoint (for subjects newly randomized in Part 2): Worsening of steatohepatitis (no worsening of lobular inflammation or hepatocellular ballooning grade 2005) on liver histology at Month 12 relative to the Screening biopsy	NAFLD activity score (NAS) components based on Screening biopsy.	Screening biopsy result before Baseline visit and before subject is randomized in Part 2

[redacted]

[redacted]

5.1.1.3.1 Endpoint Domain 1

The testing hierarchy is as follows for the Part 1 analyses:

- Global: Part 1 testing hierarchy includes 2 tests, one for the primary endpoint and the other for the key secondary endpoint, which are both binary endpoints. The key secondary endpoint is tested only when the primary endpoint result is significant. These surrogate endpoints will be tested at the 0.0012 (2-sided) alpha level to manifest strong evidence from a single confirmatory study, and 0.048 (2-sided) level for study success.
- Region-specific Europe: Part 1 testing is only for the region-specific primary endpoint, a binary endpoint. The surrogate endpoint will be tested at the 0.0012 (2-sided) alpha level to manifest strong evidence from a single confirmatory study, and 0.048 (2-sided) level for study success.

The general description of the domains of clinically-related endpoints is stated below.

Table 5-14 Domain 1 Analyses

Type of Endpoint	Description	Timing	Methodology
Part 1: Primary efficacy endpoint (binary)	For the primary efficacy analysis in Part 1, using the mITT analysis set, any available liver biopsy after baseline will be used as the Month 12 biopsy, no matter when it was obtained relative to the first dose of study drug. Subjects who do not have an evaluable liver biopsy at both Screening and Month 12 will be included in the analyses as nonresponders. Subjects with multiple liver biopsies will have the evaluable biopsy closest to the Month 12 visit included in the analysis. The proportion of subjects who meet the primary efficacy endpoint in Part 1, and the proportion who meet each component of the primary efficacy endpoint, will be summarized by treatment group. Two-sided 95% CIs for the proportion of subjects who meet the primary efficacy endpoint, calculated using Wilson's method for CI, will be reported. Analysis of the primary and key secondary endpoints of Part 1 will use the Cochran-Mantel-Haenszel (CMH) test, stratified by true fibrosis stage (2, 3) and T2DM status (presence, absence) at baseline, for a total of 4 strata, to compare the rates in the 2 randomized treatment arms.	Screening, Month 12 The primary analysis of Part 1 will occur when approximately 1200 randomized subjects have been followed for at least 12 months. Subjects who will contribute to the primary analysis of Part 1 will be identified before unblinding of any subjects for Part 1 analysis; subjects who will not contribute to the primary analysis of Part 1 will be identified by date of randomization. Individual subjects' treatment group assignments will not be disseminated to the sites until the end of Part 2 to allow for continued blinded assessment during Part 2.	Responder Two-sided 95% CIs for responder using Wilson's method for CI. Cochran-Mantel-Haenszel (CMH) test, stratified by true fibrosis stage (2, 3) and T2DM status (presence, absence) at baseline
Part 1: Primary efficacy endpoint (sensitivity analyses 1)	The primary efficacy endpoint will be performed using the ITT and PP1 populations as sensitivity analyses.	Same as above	Cochran-Mantel-Haenszel (CMH) test, stratified by actual fibrosis stage (2, 3) and T2DM status (presence, absence) at baseline
Part 1: Primary efficacy endpoint (sensitivity analysis 2)	Logistic regression, with fibrosis stage, presence of T2DM, study site as predictors and baseline covariates (to be listed out later) using the mITT analysis set.	Same as above	Logistic regression, with covariates for the stratification factors
Part 1: Primary efficacy endpoint (sensitivity analysis 3)	An assessment of efficacy will also be made using multiple imputation, for scenarios of both assuming missing at random and missing not at random, for the CMH test. Refer to Section 5.1.1.1.5.	Same as above	Multiple imputation

Type of Endpoint	Description	Timing	Methodology
Part 1: Primary efficacy endpoint (subgroup analysis)	Refer to Section 5.1.1.5.	Same as above	Cochran-Mantel-Haenszel (CMH) test, stratified by true fibrosis stage (2, 3) and T2DM status (presence, absence) at baseline
Part 1: Secondary efficacy endpoints (binary)	Analysis of other secondary efficacy endpoints of Part 1 will use similar methods to the primary endpoint analysis. Results for the mITT and PP1 analysis sets will be reported for secondary efficacy endpoints.	Same as above	Cochran-Mantel-Haenszel (CMH) test, stratified by true fibrosis stage (2, 3) and T2DM status (presence, absence) at baseline
Part 1: Secondary efficacy endpoints (continuous)	Continuous endpoints will be analyzed with analysis of covariance, with true fibrosis stage (2, 3) and T2DM status (presence, absence) at baseline) included as factors using the mITT and PP1 populations.	Same as above	CFB ANCOVA

[redacted]

Type of Endpoint	Description	Timing	Methodology
Part 2: Primary efficacy endpoint (binary responder definition)	The proportion of subjects with adjudicated clinical endpoints for Part 2, along with each component of the endpoint, will be summarized by treatment group. For the primary efficacy analysis in Part 2 using the mITT analysis set, any available liver biopsy after baseline will be used for assessment of components of the clinical endpoint that require a biopsy, no matter when obtained. Subjects who do not have an evaluable liver biopsy at both Screening and post-baseline will be included in the analyses as having an event only if they meet a component that does not require a biopsy; otherwise they will be included as not having an event and censored at the last visit. Subjects with multiple liver biopsies will be included as an event if any (one or more) biopsy meets the definition of the primary endpoint, as determined by the adjudication committee. Two-sided 95% CIs for the proportion of subjects who meet the primary efficacy endpoint, calculated using Wilson's method for CI, will be reported.	The primary analysis for Part 2 will occur when adjudicated events have been accrued in approximately 367 unique subjects. Because many events may be observed from the Month 60 liver biopsy, and given that all reported events will need to undergo independent adjudication, it is possible that a different number of subjects with events will be included in the analysis. All subjects and events in the database at the time of Part 2 database lock and unblinding will be included in the primary analysis of Part 2. If multiple events are observed in a single subject, only the first observed event will be used in the analysis.	Responder Two-sided 95% CIs for responder using Wilson's method for CI.
Part 2: Primary efficacy endpoint (time-to-event)	Time-to-event analyses will be performed using the Kaplan-Meier product-limit method with tabular and graphical presentation for the cumulative probability of event rate by treatment group. Subjects will be censored for analysis at the last recorded study visit. Analysis of the primary endpoint of Part 2 will use the logrank test, stratified by true fibrosis stage (2, 3) and T2DM status (presence, absence) at baseline, for a total of 4 strata, to compare the rates in the 2 randomized treatment arms. If multiple adjudicated events are observed in a single subject, only the first observed event will be used in the analysis. The analysis will use the mITT analysis set, with supportive analyses reported for the	Same as above	Time-to-event KM figure

Type of Endpoint	Description	Timing	Methodology
Part 2 (Global/Region-specific Europe): Testing (primary endpoint) (time-to-event)	ITT and PP2 analysis sets. Subjects who do not have an adjudicated event will be censored for analysis at the last recorded study visit. The p-value for the test of no difference between treatment arms (hazard ratio=1) will be presented as well as the hazard ratio from a proportional hazards model (CVC divided by placebo) and corresponding 95% CI. Global/Region-specific Europe: To manifest strong evidence from a single confirmatory study, the hypothesis test of clinical endpoint will be performed at a significance level of 0.00125 (2-sided) if the test of the surrogate endpoint is successful and 0.00005 (2-sided) otherwise. For study success, the hypothesis test of clinical endpoint will be performed at a significance level of 0.05 (2-sided) if the test of the surrogate endpoint is successful and 0.002 (2-sided) otherwise.		
Part 2: Primary efficacy endpoint (sensitivity analyses 1)	The primary endpoint will be performed using the ITT and PP1 populations as sensitivity analyses.	Same as above	Same as above
Part 2: Primary efficacy endpoint (sensitivity analysis 2)	An assessment of primary endpoint will also be made using multiple imputation with informative censoring imputation model. Refer to Section 5.1.1.1.5.	Same as above	Multiple imputation with informative censoring imputation model
Part 2: Primary efficacy endpoint (sensitivity analysis 3)	Analysis of the primary endpoints for Part 2 will also use the Cochran-Mantel-Haenszel (CMH) test, stratified by true fibrosis stage (2, 3) and T2DM status (presence, absence) at baseline, for a total of 4 strata, to compare the event rates in the 2 randomized tretment arms.	Same as above	Cochran-Mantel-Haenszel (CMH) test, stratified by true fibrosis stage (2, 3) and T2DM status (presence, absence) at baseline
Part 2: Primary efficacy endpoint (sensitivity analysis 4)	An assessment of primary endpoint will also be made using imputation of subjects censored without an event before 5 years of follow-up	Same as above	Time-to-event Logrank test

Type of Endpoint	Description	Timing	Methodology
Part 2: Secondary efficacy endpoints (binary)	Analysis of secondary efficacy endpoints of Part 2 will use similar methods to the primary endpoint analysis in Part 2. The primary endpoint in Part 1, improvement in fibrosis by at least one stage and no worsening of steatohepatitis at Month 12, will be reported for subjects not in the Part 1 analysis (subjects newly randomized in Part 2), and for all subjects in the study (Month 12 results from Part 1 and Part 2 combined). Results for the mITT and PP2 analysis sets will be reported for the secondary endpoints in Part 2. Binary endpoints will be analyzed with the CMH procedure, stratified by true fibrosis stage (2, 3) and T2DM status (presence, absence) at baseline.	Same as above	Cochran-Mantel-Haenszel (CMH) test, stratified by true fibrosis stage (2, 3) and T2DM status (presence, absence) at baseline

[redacted]

[redacted]

5.1.1.3.3 Multiple Comparisons Procedure for Primary and Secondary Endpoints

The significance level of both 0.00125 (2-sided) and 0.05 (2-sided) will be applied to hypothesis testing in this study, with the former to manifest strong evidence from a single confirmatory study and the latter for study success. Alpha splitting will be applied for multiplicity adjustment. The above applies for both global and EU-specific analyses.

To manifest strong evidence from a single confirmatory study, the familywise type I error rate (between the surrogate endpoint and the clinical outcome endpoint) will be controlled at the 0.00125 level by initial assignment of 0.0012 to the surrogate endpoint and 0.00005 to the clinical outcome endpoint, and the clinical outcome endpoint will be tested at the 0.00125 level if the tests of the surrogate endpoints (i.e., primary and key secondary) are successful. The testing procedure will be the same for study success, except the familywise type I error rate will be controlled at the 0.05 level instead of 0.00125 level, in which an initial alpha level of 0.048 to the surrogate endpoint and of 0.002 to the clinical outcome endpoint will be allocated.

5.1.1.4 Safety Analyses

Safety analyses will be based on the Safety Population.

No statistical hypotheses are prespecified for safety endpoints (e.g., incidence of AEs, including MACE and T2DM, clinical laboratory evaluations, vital signs, ECGs, biomarkers). Analyses will be primarily descriptive, with any p-values post-hoc and any comparative conclusions requiring confirmation.

Baseline assessments for applicable safety endpoints defined as follows:

Table 5-15 Safety Endpoint Baseline Definitions

Parameter	Description	Timing
Clinical laboratory evaluations, Vital signs, ECGs	Baseline assessment will be the last available value before the first dose of study drug for each parameter	Last non-missing assessment before first dose of study drug

5.1.1.4.1 Study Treatment Exposure and Compliance

Study treatment exposure and compliance will be summarized in total and by treatment group for the Safety Population as follows:

Table 5-16 Study Treatment Summaries

Parameter	Description	Timing	Methodology
Study treatment exposure (days)	(Treatment end date – treatment start date + 1). Number of days of treatment interruption will be subtracted from the study treatment exposure calculation.	Treatment Period	Continuous descriptive statistics
Study treatment compliance (%)	Summary by overall $100 \times \frac{\text{Number of tablets actually taken}}{\text{Number of tablets expected to be taken at the prescribed rate of 1 tablet per day (i.e., treatment end date – treatment start date + 1)}}$	Treatment Period	Continuous descriptive statistics
Categorical study treatment compliance (%)	Summary by overall • < 80% • ≥ 80 and ≤ 100% • >100 and ≤ 120% • >120%	Treatment Period	Categorical descriptive statistics

5.1.1.4.2 Adverse Events

The following adverse event (AE) terms are defined:

Table 5-17 AE Terms

Term	Description
Treatment-emergent	An event that initially occurs on or after the treatment start date, where: • Treatment start date ≤ event start date ≤ (treatment end date + 30 or last follow up visit, whichever comes later) • An exacerbation of a chronic or intermittent pre-existing condition, including either an increase in frequency and/or intensity of the condition on or after the date of the first dose of study drug
Subsequent	Any event reported that occurs or worsens after last intake of study drug + 30 days.

AEs-reported as occurring after the Screening Visit, will be coded using MedDRA version 22.1 or newer. Unique subjects reporting AEs in the following AE categories will be summarized by treatment group and in total (Part 1 and Part 2) for the Safety Population as follows:

Table 5-18 AE Summaries

Parameter	Description	Timing	Methodology

Overall summary	Overall summary only for the following categories:
	• TEAEs
	  ○ Grade 2 or above
	  ○ Grade 3 or above
	• Treatment-related TEAEs
	  ○ Grade 2 or above
	  ○ Grade 3 or above
	• TESAEs
	• Treatment-related TESAEs
	• TEAEs leading to study drug discontinuation
	• Deaths
		From treatment start date until 30 days after treatment end date or last follow up visit, whichever comes later	Categorical descriptive statistics

TEAEs	Overall summary and by SOC and PT	From treatment start date until 30 days after treatment end date or last follow up visit, whichever comes later	Categorical descriptive statistics
TEAEs by severity	Overall summary and by SOC, PT, and severity • Subjects categorized overall and within each SOC and PT for the most intense occurrence • If more than one adverse event with the same preferred term was reported, then the adverse event with the greatest severity will be used.	From treatment start date until 30 days after treatment end date or last follow up visit, whichever comes later	Categorical descriptive statistics
Treatment-related TEAEs	Overall summary and by SOC and PT	From treatment start date until 30 days after treatment end date or last follow up visit, whichever comes later	Categorical descriptive statistics
Treatment-related TEAEs by severity	Overall summary and by SOC, PT, and severity • Subjects categorized overall and within each SOC and PT for the most intense occurrence If more than one adverse event with the same preferred term was reported, then the adverse event with the greatest severity will be used.	From treatment start date until 30 days after treatment end date or last follow up visit, whichever comes later	Categorical descriptive statistics
TESAEs	Overall summary and by SOC and PT	From treatment start date until 30 days after treatment end date or last follow up visit, whichever comes later	Categorical descriptive statistics

Parameter	Description	Timing	Methodology
Treatment-related TESAEs	Overall summary and by SOC and PT	From treatment start date until 30 days after treatment end date or last follow up visit, whichever comes later	Categorical descriptive statistics
TEAEs leading to study drug discontinuation	Overall summary and by PT	From treatment start date until 30 days after treatment end date or last follow up visit, whichever comes later	Categorical descriptive statistics
MACE and new onset of T2DM	Overall summary and by SOC and PT • Includes PTs, high level terms (HLTs), and/or standardized MedDRA queries (SMQs) defined in Section 6.6.1.3	From treatment start date until 30 days after treatment end date or last follow up visit, whichever comes later	Categorical descriptive statistics
TEAEs by subgroups of age, sex, race, and ethnicity	Overall summary and by SOC and PT	From treatment start date until 30 days after treatment end date or last follow up visit, whichever comes later	Categorical descriptive statistics

SOCs will be sorted alphabetically; PTs will be sorted in descending frequency.
All AEs (including non-treatment-emergent events) recorded on the eCRF will be listed for individual subjects showing both verbatim and preferred terms.

5.1.1.4.3 Clinical Laboratory Assessments

Clinical laboratory assessments will be summarized by treatment group for the Safety Population as follows:

Table 5-19 Clinical Laboratory Summaries

Type of Endpoint	Description	Timing	Methodology
Laboratory values over time	Summary by laboratory category and parameter in SI units and analysis visit • Parameters specified in Section 6.6.2.2	Part 1: Screening, Baseline (Day 1), Months 1, 3, 6, 9, 12, and 30-Day Follow-up, Early DC[1] Part 2 (subjects newly randomized in Part 2): Screening, Baseline, Months 1, 6, 12 Part 2 (all subjects (continuing from Part 1 and newly randomized in Part 2): Every 6 months from Month 18 through Month 96 End of Study Part 1 and Part 2: Follow-up, Early Discontinuation Visit[1]	CFB descriptive statistics
Laboratory abnormalities by toxicity grade	Summary by laboratory category and parameter • Grade 1, Grade 2, Grade 3 and Grade 4 categories based on NCI CTCAE version 4.03 • Parameters specified in Table 6-8 • Include all post-baseline combined for counts of toxicity by grade. Only treatment-emergent laboratory abnormalities should be summarized in laboratory abnormalities by grade – must have a toxicity that is a higher grade than at baseline to be included.	Part 1: Screening, Baseline (Day 1), Months 1, 3, 6, 9, 12, and 30-Day Follow-up, Early DC[1] Part 2 (subjects newly randomized in Part 2): Screening, Baseline, Months 1, 6, 12 Part 2 (all subjects (continuing from Part 1 and newly randomized in Part 2): Every 6 months from Month 18 through Month 96 End of Study Part 1 and Part 2: Follow-up, Early Discontinuation Visit[1]	Categorical descriptive statistics

[1] Based on nominal visits defined in the schedule of assessments.

5.1.1.4.3.1 Liver Biochemistry

Elevations of ALP, ALT, AST, or bilirubin, or confirmed ALP, ALT, AST, or bilirubin elevations (with or without liver-related clinical symptoms) will be summarized by treatment group for the Safety Population as follows:

Table 5-20 Liver Biochemistry Summaries

Type of Endpoint	Description	Timing	Methodology
ALP, ALT, AST, or bilirubin elevations with or without liver-related clinical symptoms	ALP, ALT, AST, or bilirubin elevations with or without liver-related clinical symptoms (with or without confirmatory measurement), based on protocol Table 11-1	Treatment Period + 30 days after treatment end date	Categorical descriptive statistics
Confirmed ALP, ALT, AST, or bilirubin elevations with or without liver-related clinical symptoms	Confirmation of ALP, ALT, AST, or bilirubin elevations based on protocol Table 11-1. Confirmatory measurements are obtained <3 days, 3-5 days, or >5 days; adjudication decisions are classified as unlikely, possible, and probable DILI (drug-induced liver injury).	Treatment Period + 30 days after treatment end date	Categorical descriptive statistics

ALP = alkaline phosphatase; ALT = alanine aminotransferase; AST = aspartate aminotransferase

5.1.1.4.4 Vital Signs

Vital signs will be summarized by treatment group and in total for the Safety Population as follows:

Table 5-21 Vital Signs Summaries

Type of Endpoint	Description	Timing	Methodology
Vital signs over time	Summary by parameter and analysis visit • Parameters specified in Section 6.6.3.2	Part 1: Screening, Baseline, Treatment Period (Months 1, 3, 6, 9, 12), 30-Day FU, Early DC Part 2 (subjects newly randomized): Screening, Baseline, Treatment (Months 1, 6, 12) Part 2 (all subjects – continuing from Part 1 and newly randomized in Part 2): Through to Month 60 biopsy, Post biopsy, EOS, 1-Month FU, Early DC[1]	CFB descriptive statistics

[1] Based on nominal visits defined in the schedule of assessments.

5.1.1.4.5 Electrocardiograms

Electrocardiograms (ECGs) will be summarized by treatment group and in total for the Safety Population as follows:

Table 5-22 ECG Summaries

Type of Endpoint	Description	Timing	Methodology
ECG parameter values over time	Summary by parameter and analysis visit • Parameters specified in Section 6.6.4.3	Part 1: Baseline, Treatment Period: Month 12 Part 2: (subjects newly randomized): Baseline, Treatment Period: Month 12 Part 2 (all subjects –continuing from Part 1 and newly randomized in Part 2): Through to Month 60 Biopsy, and Post biopsy[1]	CFB descriptive statistics

[1] Based on nominal visits defined in the schedule of assessments.

5.1.1.4.6 Other Analyses

Not applicable.

5.1.1.5 Subgroup Analyses

In Part 1, subgroup analyses for the primary efficacy endpoint using the CMH test will be reported for selected subgroups to be defined in this section (table below). These analyses will place emphasis on the comparison of CVC to placebo within subgroups and assessing the consistency of magnitude of the treatment effect across subgroups. The odds ratios from the CMH analysis will be reported with associated 2-sided 95% CIs.

In addition, subgroup analyses will be conducted for the key secondary endpoint and secondary endpoints for NASH CRN fibrosis stage (2 vs. 3) and for T2DM at baseline (yes, no) (based upon the data entries in the electronic database).

Subgroup results (treatment effect estimate and two-sided 95% CIs) will be presented on a Forest plot also displaying the overall treatment effect and its 95% CI. The emphasis will be on assessing the consistency of the treatment effect across subgroups, i.e., overlapping CIs indicate consistent treatment effects.

The table below also displays the rationale for pre-specifying the specific subgroup i.e. regulatory requirements to assess consistency of effect for major demographic factors (e.g., age, sex, and race or ethnicity), [redacted]

For some subgroups, detailed information with many levels are collected (e.g., race/ethnicity); frequencies of subjects for each separate subgroup level will be reported as part of the demographic and baseline characteristics summaries. However, to provide an assessment of the consistency of efficacy and safety results across major subgroup categories, some levels will be combined into 3 to 4 major categories for each subgroup. This will ensure that the number of subjects per category is sufficient to perform meaningful subgroup analyses.

The list of subgroups may be modified depending on the recruitment to the trial to ensure the number of subjects is sufficient in the subgroups to be assessed. This will be determined before unblinding.

Table 5-23 Subgroups

Subgroup	Categories	Comments
Age at baseline (years)	< 65, ≥ 65	
Sex	Male, Female	
Race	White, Non-white	
NASH CRN fibrosis stage at baseline	2, 3	Data from Phase 2b to be confirmed or refuted. To be based on the true data (e.g., as reported in the CRF rather than in IxRS)

[redacted]

Subgroup	Categories	Comments
Geographic region	USA only EU only USA/Canada vs. Rest of World (ROW) ROW: • EU • Australia/New Zealand • Middle East/Africa • Asia Pacific (excluding Australia/New Zealand) • Latin America (Mexico, Central and South Americas) (Note: a full list of countries will be compiled after enrolment of Part 1 is complete and before unblinding of Part 1, and each will be assigned to a geographic region. Additional countries added after unblinding of Part 1, if any, will be assigned to a geographic region for analysis of Part 2.)	

[redacted]

[redacted]

[redacted]

[redacted]

5.1.1.6 Interim Analyses

5.1.1.6.1 Part 1 Analyses

The Part 1 analyses will occur when approximately 1200 randomized subjects have been followed through the Month 12 visit assessment. Subjects who will contribute to the primary analysis of Part 1 will be identified before unblinding of any subjects for Part 1 analysis.

To perform the Part I analyses, an unblinded team will be identified whose members will be unblinded to the treatment information at the individual subject level. The unblinded team members will no longer be involved in any ongoing daily conduct of the blinded study in Part 2. Other blinded personnel will assume or continue their responsibilities until the end of Part 2.

or study team members who will not continue activities subsequent to the Part 1 lock. Individual subjects' treatment group assignments will not be disseminated to the sites until the end of Part 2 to allow for continued blinded assessment during Part 2.

5.1.1.6.2 Data and Safety Monitoring Board

An independent DSMB will be formed to review ongoing data from this study. The DSMB will be composed of members who have medical expertise in liver disease and at least one statistician. The DSMB will be empowered to recommend changes to the protocol to ensure the safety of subjects in the study but will not be empowered to recommend stopping or changing the study due to accumulating efficacy data, other than at the planned futility analysis in Part 2 of the study.

Safety data will be reviewed by the DSMB. The first review will occur after the first subject has been followed for at least 6 months. Subsequent reviews will be at least quarterly, at a schedule agreed by the sponsor and the members of the board per the DSMB charter. This data review will not result in any adjustments to alpha because there will not be a chance to stop the study early for a conclusion of efficacy.

For Part 2 of the study, one unblinded assessment of futility may be performed by the DSMB. This may occur after the last randomized subject has been followed for at least 2 years and at least half of the expected events have been observed. If the conditional power exceeds 5%, the study will continue. The study will be stopped only for a conclusion of futility. This data review will not result in any adjustments to alpha because there will not be a chance to stop the study early for a conclusion of efficacy.

No investigator or subject will be unblinded to support DSMB review. No one involved in data review or subject care will be unblinded.

5.1.1.6.3 Adjudication Committee

An independent adjudication committee will be formed to review all events which are potentially components of the primary composite endpoint for Part 2 of the study. Only events confirmed by the adjudication committee will be included in the analysis of the primary endpoint of Part 2; the time to onset will be determined as the date on which the event occurred according to the adjudication committee.

5.1.2 Determination of Sample Size

Part 1:

The sample size for the primary endpoint of Part 1 is based on the primary binary endpoint at the end of Month 12 comparing treatment with CVC versus placebo. The planned sample size of 1200 subjects (800 in treatment Arm A and 400 in treatment Arm B) for Part 1 of this study is expected to provide 84% power to demonstrate strong evidence with a single study (2-sided alpha level of 0.0012), assuming a 15% response rate for the placebo arm and a 25% response rate for CVC according to the results from the Phase 2b Study 652-2-203 (CENTAUR).

The planned 800 subjects in treatment Arm A and 400 subjects in treatment Arm B is expected to provide 97% power to demonstrate strong evidence with a single study (2-sided alpha level of 0.0012) in the key secondary endpoint, assuming a 2.2% response rate for the placebo arm and an 8.6% response rate for CVC according to the results from the Phase 2b Study 652-2-203 (CENTAUR).

PASS 8.0 is used to calculate the sample size.

These response rates reflect the primary analysis, in which subjects missing the post-baseline liver biopsy will be included as nonresponders.

The anticipated proportion of subjects with missing Month 12 liver biopsies, as a result of either premature subject discontinuation or nonevaluable liver biopsy results (i.e., biopsy sample deemed inadequate for evaluation of efficacy endpoints by an independent central pathologist), is estimated to be 15% of subjects at Month 12, compared to approximately 13% of missing post-treatment liver biopsies in the Phase 2b study.

Part 2:

The sample size for the primary endpoint analysis in Part 2 is based on the estimated event-free survival rate of 80% for the placebo group and detection of a hazard ratio of 0.62 by the anticipated end of the study at Month 60 (corresponding to a median event-free survival time of approximately 15 years for placebo and 25 years for CVC). A total of 2000 subjects (2:1 randomization ratio between CVC and placebo) enrolled approximately uniformly over 2 years for an overall study duration of approximately 8 years (2 years of accrual period plus 5 to 6 years of follow-up) will lead to about 367 events, after accounting for an overall dropout rate of 20%. With these events, there will be 85% power to demonstrate strong evidence of superiority of CVC over placebo (at a 2-sided 0.00125 significance level, for a single registration study), and 99% power to test the superiority of CVC over placebo (at the 2-sided test 0.05 significance level for a registration study).

EAST 6.4 is used for the calculation.

5.2 Changes in the Conduct of the Study or Planned Analyses

5.2.1 Changes in the Conduct of the Study

Approximately 1200 randomized participants were planned for the Part 1 analyses. During the actual study execution during the screening period for Part 1, 1293 participants were randomized into Part 1. This affects the newly randomized subject enrollment numbers anticipated for Part 2, as the number newly randomized in Part 2 will be decreased from approximately 800 to approximately 707. However, the total number randomized remains the same:

Arm	Newly Randomized Subjects in Part 2 N	Total Subjects (Randomized in Part 1 or Part 2) N	Treatment
A	472	1334	CVC 150 mg, once daily
B	235	666	Placebo, once daily

5.2.2 Changes to Analyses Prior to Database Lock

[redacted]

- PRO-C3 was added as a liver fibrosis measure (Section 4.2). It will be analyzed with descriptive statistics over time, including change over time and value over time.
- The Enhanced Liver Fibrosis (ELF™) Score was added as another liver fibrosis index (Section 4.2). It will be analyzed with descriptive statistics over time, including change over time, value over time, and correlation with other fibrosis indices and histological indices.
- Subgroup analyses were updated (Section 5.1.1.5). Subgroup analyses were added for NASH CRN fibrosis stage and for T2DM for the key secondary and secondary endpoints. In addition, medication use and weight loss subgroups were added, while Framingham score and smoking status were removed.

6. Data Handling and Analysis Conventions

6.1 Study Treatment Conventions

Treatment assignment will be done using permuted block randomization via IxRS stratified by NASH CRN fibrosis stage (2 or 3) and presence or absence of documented T2DM (yes or no). In case of discordance of the assignment of randomization strata between IxRS and EDC, the true fibrosis stage and T2DM status at baseline will be used for the analysis.

6.1.1 Analysis Days

Treatment days are defined as follows:

Table 6-1 Analysis Day Definitions

Term	Description
Treatment Day	Relative to treatment start date If analysis date ≥ treatment start date: • Day = analysis date – treatment start date + 1 o Day 1 = treatment start date If analysis date < treatment start date: • Day = analysis date – treatment start date o Day -1 = day before treatment start date o There is no Day 0

6.1.2 Missing/Incomplete Treatment End Date

If the investigator is unable to provide the treatment end date, the last available dosing record date will be used as treatment end date.

6.2 Analysis Visit Windows

6.2.1 Efficacy

The analysis visit windows for efficacy endpoints, including liver biopsy, are defined as follows:

Table 6-2 Efficacy Analysis Visit Definitions

Analysis Phase	Analysis Visit (Derived)	Study Visit (eCRF)	Window
Part 1 & Part 2			
Pretreatment	Baseline	Day 1	Treatment Day ≤ 1, pre-randomization in Part Y
Treatment	Month 12	Day 360	Treatment Day [316, 405], post-randomization in Part 2 Note: For the primary efficacy analysis in Part 1 using the mITT analysis set, any available liver biopsy after randomization will be used for the Month 12 biopsy, no matter when it was obtained relative to the first dose of study drug. Subjects with multiple liver biopsies will have the evaluable biopsy closest to the Month 12 visit included in the analysis.
	Month 60	Day 1800	Treatment Day [1756, 1845], post-dose in Part 2
	End of study	Final or termination visit	Last non-missing assessment during the Treatment Period in Part Y / between Treatment Day 1 in Part Y, post-dose, and Off-treatment Day 1, inclusive, where Y=1, 2.

The following general conventions for repeated or unscheduled assessments will apply unless otherwise specified:

- For scheduled and unscheduled assessments, if more than one assessment is measured within the same visit window, then the record with the latest visit date (from baseline) will be used for that visit window.
- All assessments will be included in respective listings.

6.2.2 Safety

The analysis visit windows for safety endpoints, including laboratory assessments, are defined as follows:

Table 6-3 Safety Analysis Visit Definitions

Analysis Phase	Analysis Visit (Derived)	Scheduled Study Visit (eCRF)	Window
Part 1 & Part 2			
Treatment	Month 1	Day 30	Treatment Day [2, 60]
	Month 3	Day 90	Treatment Day [61, 135]
	Month 6	Day 180	Treatment Day [136, 225]
	Month 9	Day 270	Treatment Day [226, 315]
	Month 12	Day 360	Treatment Day [316, 405]
	Month X (every 6 months)	Day X*30	Treatment Day [X*30-44, X*30+45], where X=18, 24, 30, …, 84, 90, 96
	30-Day FU	Day X*30 +30	Treatment Day [X*30-44, X*30+45] +30, where X=18, 24, 30, …, 84, 90, 96
	End of study	Final or termination visit	Last non-missing assessment after treatment start date in Part Y / between Treatment Day 2 in Part Y and Off-treatment Day 30, inclusive, where Y=1,2.

The following general conventions for repeated or unscheduled assessments will apply unless otherwise specified:

- For scheduled and unscheduled assessments, if more than one assessment is measured within the same visit window, then the record with the latest visit date (from baseline) will be used for that visit window.
- All assessments will be included in respective listings.

For plasma and serum samples for storage, samples at visits beyond baseline (Visit 2) can be used for checking polymorphisms, genotype status, and pharmacogenetic biobanking.

6.3 Site Pooling

Refer to Section 5.1.1.1.5.

6.4 Missing/Incomplete Date Conventions

Dates may be imputed with year, month, and day values under certain scenarios:

Table 6-4 Imputation Scenarios

Scenario	Complete			Imputable
	Year	Month	Day	
1	Yes	Yes	Yes	Complete
2	Yes	Yes	—	Yes
3	Yes	—	Yes	No[1]
4	Yes	—	—	Yes
5	—	Yes	Yes	No[1]
6	—	Yes	—	No[1]
7	—	—	Yes	No[1]
8	—	—	—	Yes

[1] Not allowed per database design.

Dates will be imputed initially toward a specified target date for imputable scenarios 2, 4, and 8, and adjusted against the latest reasonable dates. The initial imputed date is determined by the following algorithm:

Table 6-5 Initial Imputed Date Algorithm

Available Year (YYYY)	Available Month (MM)			
	Missing	< Target Month	= Target Month	> Target Month
Missing	Target Date	—		
< Target Year	YYYY-12-31	YYYY-MM-LD		
= Target Year	Target Date	YYYY-MM-LD	Target Date	YYYY-MM-01
> Target Year	YYYY-01-01	YYYY-MM-01		

YYYY = available start date year; MM = available start date month; LD = last day of the month.

6.4.1 Missing/Incomplete AE Start Date

AE start dates will be imputed as the minimum of the following:

- Initial imputed date, where target date = Treatment start date
- Complete end date

6.4.2 Missing/Incomplete Medication Start Date

Medication start dates will be imputed as the minimum of the following:

- Initial imputed date, where target date = Treatment start date – 1
- Complete end date

6.4.3 Missing/Incomplete AE/Medication End Date

AE and medication end dates will be imputed as the minimum of the following:

- Initial imputed date, where target date = Treatment end date + 30
- Death date

6.5 Efficacy Endpoint Conventions

No worsening of lobular inflammation or hepatocellular ballooning grade (2005) is defined as change from baseline ≤ 0, where change from baseline = follow-up value minus baseline value.

6.6 Safety Endpoint Conventions

6.6.1 Adverse Events

6.6.1.1 Missing Intensity or Relationship

If the investigator is unable to provide the actual values, the following imputations will be applied for tabular summaries (listings will present data as missing):

Table 6-6 Missing AE Intensity and Relationship Imputation Algorithms

Missing Value	Imputation	Timing
Intensity	Mild	Pretreatment Period
	Severe	Treatment Period
Relationship	—	Pretreatment Period
	Related	Treatment Period

6.6.1.2 Possible Distant Spread of Toxin (PDSOT)

Not applicable.

6.6.1.3 AE Group of Interest

Elevations in biochemistry that are associated with liver injury have been identified as potentially important risks and are considered AESI for the study drug in this protocol. Biochemical criteria for suspected DILI are defined in protocol Section 11.3.6.4. Cases of suspected DILI as defined by the prespecified criteria in protocol Section 11.3.6.4 should be reported to the sponsor within 24 hours on the AESI form and will be adjudicated by a hepatologist with expertise in DILI and reviewed by the DSMB.

6.6.1.4 Major Adverse Cardiovascular Events (MACE)

MACE (major adverse cardiovascular events) will be evaluated based on pre-defined diagnostic criteria in accordance with current cardiovascular guidelines for MACE[3-5] which consists of 4 endpoints, or MACE-plus:

1. cardiovascular death,
2. non-fatal myocardial infarction,
3. non-fatal stroke, and
4. hospitalization for cardiovascular causes (including unstable angina).

Cases of suspected MACE will be identified by carrying out surveillance for potential MACE that may become cases of confirmed CV events during the study by (1) reviewing reports of all Serious Adverse Events (SAEs), and (2) searching the database of Treatment-Emergent Adverse Events (TEAEs) for Standardized MedDRA Query (SMQ) terms for the following preferred terms: myocardial infarction, other ischemic heart disease, heart failure, ischemic central nervous system vascular conditions, hemorrhagic central nervous system vascular conditions. The preferred terms could include derivatives related to the above.

Those reports will be assigned to any of the 4-point MACE described as follows:

1. **Cardiovascular Death**
   - Sudden cardiac death
   - Fatal myocardial infarction
   - Fatal stroke
   - Other fatal cardiovascular events (e.g., death due to heart failure, arrhythmia, pulmonary embolism, aortic dissection/rupture, etc.).
2. **Non-fatal Myocardial Infarction**
   - ST elevation myocardial infarction (STEMI)
   - Non-ST elevation myocardial infarction (NSTEMI)
3. **Non-fatal Stroke**
   - Ischemic/embolic
   - Hemorrhagic
   - Undetermined
4. **Hospitalization for Cardiovascular Causes**
   - Acute coronary syndrome (ACS) or unstable angina requiring hospitalization (including urgent revascularization procedure)
   - Heart failure requiring hospitalization

- Arrythmia (non-fatal) not associated with ischemia events, i.e., treatment emergent atrial fibrillation, supraventricular tachycardia, ventricular tachycardia, high-grade AV-block or ventricular fibrillation

All preferred terms and cases will be identified prior to database lock.

The MedDRA SMQs are below:

Broad Term SMQ	Narrow Term SMQ
Cardiac arrhythmias (SMQ)	Arrhythmia related investigations, signs and symptoms (SMQ)
	Bradyarrhythmia terms, nonspecific (SMQ)
	Conduction defects (SMQ)
	Disorders of sinus node function (SMQ)
	Cardiac arrhythmia terms, nonspecific (SMQ)
	Supraventricular tachyarrhythmias (SMQ)
	Tachyarrhythmia terms, nonspecific (SMQ)
	Ventricular tachyarrhythmias (SMQ)
Cardiac failure (SMQ)	Cardiac failure (SMQ)
Central nervous system vascular disorders (SMQ)	Conditions associated with central nervous system haemorrhages and cerebrovascular accidents (SMQ)
	Haemorrhagic central nervous system vascular conditions (SMQ)
	Ischaemic central nervous system vascular conditions (SMQ)
	Central nervous system vascular disorders, not specified as haemorrhagic or ischaemic (SMQ)
Ischaemic heart disease (SMQ)	Myocardial infarction (SMQ)
	Other ischaemic heart disease (SMQ)

6.6.1.5 Type 2 Diabetes Mellitus (T2DM)

New-onset T2DM will be identified using the MedDRA preferred terms (PTs). TEAEs will be compared to the medical history to confirm new onset versus an existing condition.

The MedDRA PTs and PT codes are below:

Preferred Term	PT Code
Insulin-requiring type 2 diabetes mellitus	10053247
Type 2 diabetes mellitus	10067585

6.6.2 Clinical Laboratory Assessments

6.6.2.1 Potentially Clinically Significant Criteria

Any laboratory abnormalities deemed clinically significant by the investigator must be reported as an AE. A clinically significant abnormality is a confirmed abnormality (by repeat testing) that is changed sufficiently from Baseline so that in the judgment of the investigator a change in management is warranted. Any laboratory test showing abnormal results (including those recorded as AEs) that are believed to be possibly/probably related to study drug treatment will be repeated weekly (or as often as deemed prudent by the investigator) until the abnormality is resolved, returns to Baseline, or is otherwise explained. Laboratory assessments values meeting *any* of the following PCS low or PCS high criteria will be categorized as PCS:

Table 6-7 Clinical Laboratory PCS Criteria

Category	Parameter	SI Unit	PCS Criteria	
			PCS Low	PCS High
Chemistry	Albumin	g/L	< 0.9 × LLN	> 1.1 × ULN
	Alanine aminotransferase	U/L	—	≥ 3.0 × ULN
	Alkaline phosphatase	U/L	—	≥ 2.0 × ULN
	Aspartate aminotransferase	U/L	—	≥ 3.0 × ULN
	Bilirubin, total or direct	μmol/L	—	> 1.5 × ULN
	Calcium	mmol/L	< 0.9 × LLN	> 1.1 × ULN
	Chloride	mmol/L	< 0.9 × LLN	> 1.1 × ULN
	Cholesterol	mmol/L	—	> 1.3 × ULN
	Creatinine	μmol/L	—	> 1.3 × ULN
	Glucose, fasting	mmol/L	< 0.8 × LLN	> 1.2 × ULN
	Glucose, nonfasting	mmol/L	< 0.8 × LLN	> 1.4 × ULN
	Potassium	mmol/L	< 0.9 × LLN	> 1.1 × ULN
	Protein, total	g/L	< 0.9 × LLN	> 1.1 × ULN
	Sodium	mmol/L	< 0.9 × LLN	> 1.1 × ULN
	Triglycerides	mmol/L	—	> 2.0 × ULN
	Urea nitrogen	mmol/L	—	> 1.2 × ULN
	Uric acid	μmol/L	—	> 1.2 × ULN
Hematology	Basophils, absolute cell count	10^9/L	—	> 3.0 × ULN
	Neutrophils, absolute cell count	10^9/L	< 0.8 × LLN	> 1.5 × ULN
	Hematocrit	Ratio	< 0.9 × LLN	> 1.1 × ULN
	Hemoglobin	g/L	< 0.9 × LLN	> 1.1 × ULN
	Platelet count	10^9/L	≤ 0.5 × LLN	≥ 1.5 × ULN
	Red blood cell count	10^{12}/L	< 0.9 × LLN	> 1.1 × ULN
	White blood cell count	10^9/L	≤ 0.7 × LLN	≥ 1.5 × ULN
Urinalysis	pH	—	< 0.9 × LLN	> 1.1 × ULN
	Specific gravity	—	—	> 1.1 × ULN

LLN = lower limit of normal value; ULN = upper limit of normal value; normal value provided by laboratory.
SI = Le Système International d'Unités (International System of Units).

6.6.2.2 Continuous Descriptive Parameters

The following laboratory parameters will be summarized:

Table 6-8 Clinical Descriptive Parameters

Category	Parameters
Hematology	Hematocrit, Hemoglobin, Mean corpuscular hemoglobin (MCH), Mean corpuscular hemoglobin concentration (MCHC), Mean corpuscular volume (MCV), Platelet count, Red blood cell distribution width (RDW), Red blood cell count, White blood cell count, White blood cell differential (absolute counts only): Basophils, Eosinophils, Lymphocytes, Monocytes, Neutrophils Part 1: Every visit Part 2: Subjects continuing from Part 1: every 6 months through the end of the study Newly enrolled subjects: Screening, Baseline, Months 1, 6, and 12 and every 6 months thereafter through the end of the study
Chemistry	Alkaline phosphatase (ALP), Alanine aminotransferase (ALT), Albumin, Aspartate aminotransferase (AST), Direct bilirubin, Gamma-glutamyl transferase (GGT), Total bilirubin, Amylase (reflex lipase if amylase ≥ 1.5 × ULN), Bicarbonate, Blood urea nitrogen (BUN), Calcium, Chloride, Cholesterol (total), Creatine phosphokinase (CPK), Creatinine, Glucose International normalized ratio (INR), Lactate dehydrogenase (LDH), Magnesium, Phosphorus, Potassium, Sodium, Total protein, Triglycerides, Uric acid Part 1: Every visit Part 2: Subjects continuing from Part 1: every 6 months through the end of the study Newly randomized subjects: Screening, Baseline, Months 1, 6, and 12 and every 6 months thereafter through the end of the study
Serum Hepatic Fibrosis Indices	Fibrosis 4 Score (FIB-4): Platelets, ALT, AST, Subject age Aspartate aminotransferase-to-platelet-count ratio (APRI): AST to platelet count ratio index, NAFLD Fibrosis Score (NFS): Subject age, BMI, AST, ALT, Platelets, Albumin ELF score: Subject age, BMI, AST, ALT, Platelets, Albumin PRO-C3 Part 1: Baseline, Months 6, and 12 Part 2: Subjects continuing from Part 1: every 12 months through end of study Newly randomized subjects: Screening, Baseline, Months 6 and 12, and every 12 months through end of study
Fasting Metabolic Parameters	The following parameters may be included, but are not limited to: Lipid panel: Cholesterol (total), High density lipoprotein (HDL), Low density lipoprotein (LDL), Triglycerides, Very low-density lipoprotein (VLDL), Glucose, Insulin, Hemoglobin A1c (HbA1c) Part 1: Baseline, Months 3, 6, and 12 Part 2: Subjects continuing from Part 1: every 6 months through the end of the study Newly randomized subjects: Baseline, every 6 months through the end of the Study
Biomarkers	The following parameters may be included, but are not limited to: Fibrinogen, High-sensitivity C-reactive protein (hs-CRP), Interleukin (IL)-1β, IL-6 Part 1: Baseline, Months 3, 6, and 12 Part 2: Subjects continuing from Part 1: every 6 months through the end of the study Newly randomized subjects: Baseline, every 6 months through the end of the Study

Category	Parameters
Other Tests: Part 1 and Part 2	Screening only: Ferritin, Follicle-stimulating hormone, (FSH) for postmenopausal women only Hemoglobin A1c (HbA1c), Hepatitis B surface antigen (HBsAg), Hepatitis C virus antibody (HCVAb), HCV RNA (only if HCVAb positive), HIV-1 and -2 antibodies Screening, Month 12 and annually thereafter: α-fetoprotein (AFP) All visits Urine pregnancy test for females of childbearing potential only. Positive urine pregnancy tests will be confirmed with a serum test.
Hepatic Panel Testing	Part 1 and Part 2: Baseline and Unscheduled Visits The following parameters may be included, but are not limited to: Baseline only Anti-liver cytosol type 1 (anti-LC1), Antinuclear antibody (ANA), Anti-liver/kidney microsome type 1 (anti-LKM1) antibodies, Anti-smooth muscle antibodies (anti-SMA), Immunoglobulin G (IgG), Hepatitis B core antibody (HBcAb), Hepatitis B surface antibody (HBsAb) titer Unscheduled Visits Antimitochondrial antibodies (AMA), ANA, Anti-LKM-1 type antibodies, anti-LC1 antibodies, Hepatitis A virus immunoglobulin M (HAV-Ab IgM), Hepatitis B virus surface antigen (HBsAg), Hepatitis C virus antibody (HCVAb), Hepatitis D virus antibody, Hepatitis E virus RNA by PCR, Immunoglobulin G subclass 1, anti-SMA, Alkaline phosphatase (ALP), Alanine aminotransferase (ALT), Aspartate aminotransferase (AST), Direct bilirubin, Total bilirubin

6.6.2.3 Character Values

Character values (e.g., < 5, negative) will be reviewed prior to database lock and converted to numeric values for analysis as appropriate. These conversions will be documented in the ADaM specifications.

6.6.3 Vital Signs

6.6.3.1 Potentially Clinically Significant Criteria

For vital signs, the PCS criteria are for weight change, as shown below:

- Normal: ≤7% change from baseline
- Abnormal low decrease: Decrease of >7 to ≤10% from baseline
- Abnormal low increase: Increase of >7 to ≤10% from baseline
- Abnormal high decrease: Decrease of >10% from baseline
- Abnormal high increase: Increase of >10% from baseline

6.6.3.2 Continuous Descriptive Parameters

The following vital sign parameters will be summarized:

Table 6-9 Vital Sign Descriptive Parameters

Parameters		
Systolic blood pressure (mmHg)	Diastolic blood pressure (mmHg)	Pulse rate (bpm)
Respiratory rate (breaths/min)	Temperature (°C)	Weight (kg)
BMI (kg / m^2)	Waist Circumference (cm)	Hip Circumference (cm)
Waist-to-hip Ratio		

6.6.4 Electrocardiograms

6.6.4.1 QTc Derivation

QTc Bazett (QTcB) and QTc Fridericia (QTcF) are derived as follows:

Table 6-10 ECG PCS Criteria

Parameter	Derivation if RR available	Derivation if RR unavailable
QTcB	$\frac{QT}{\text{square root of RR}}$	$\frac{QT}{\text{square root of 1/HR}}$
QTcF	$\frac{QT}{\text{cubic root of RR}}$	$\frac{QT}{\text{cubic root of 1/HR}}$

QTcB = QTc Bazett; QTcF = QTcF Fridericia.

6.6.4.2 Potentially Clinically Significant Criteria

Any abnormal findings that are considered clinically significant in the opinion of the investigator will be recorded as AEs or be captured as medical history, if already present at Screening. ECG results will be reviewed for clinically notable abnormalities according to predefined criteria. Subjects exhibiting Grade 3 or 4 PR or QT interval corrected for heart rate (QTc) interval will be summarized. Abnormalities in Fridericia's corrected QT interval (QTcF) interval, Bazett's corrected QT interval (QTcB) interval, QRS, PR, and heart rate (HR) will be summarized.

ECG results will be reviewed for clinically notable abnormalities according to the criteria described below.

For values of QTc, the categories are:

- Normal: Less than 450 msec
- Borderline: At least 450 msec and up to 480 msec

- Prolonged: More than 480 msec and up 500 msec
- Pathologically prolonged: More than 500 msec

For QTc change from baseline, the categories are:

- Normal: Decrease/no change/increase up to 30 msec
- Abnormal: Increase of more than 30 msec up to 60 msec
- Abnormal high: Increase of more than 60 msec

For QRS, the categories are:

- Abnormal low: Up to 50 msec
- Normal: More than 50 msec and less than 120 msec
- Abnormal high: At least 120 msec

For PR, the categories are:

- Normal: Less than 210 msec
- Abnormal: At least 210 msec

For HR, the categories are:

- Abnormally low HR: ≤50 bpm
- Normal HR: > 50 bpm and <120 bpm
- Abnormally high HR: ≥120 bpm

Abnormalities are the last three categories listed for QTc, the last two for QTc change from baseline, the first and last for QRS, the last for PR and the first and last for HR.

In addition, Grade 3 or 4 QT interval corrected for heart rate (QTc interval) will be derived based on the ECG data and following the NCI CTCCAE definition.

- For QT, Grade 3 is defined as QTc≥ 501 msec on at least 2 separate ECGs and Grade 4 is defined as QTc≥ 501 msec or >60 ms change from baseline and Torsade de pointes or polymorphic ventricular tachycardia or signs/symptoms of serious arrhythmia.

6.6.4.3 Continuous Descriptive Parameters

The following ECG parameters will be summarized:

Table 6-11 ECG Descriptive Parameters

Parameters		
Heart rate	QRS interval	QT interval
	PR interval	QTcB
		QTcF

QTcB = QTc Bazett; QTcF = QTc Fridericia.

6.7 Imputed Value Listing Conventions

In general, listings will present the actual partial or missing values rather than the imputed values that may be used in endpoint derivation. In instances where imputed values will be presented, imputed values will be flagged. Actual rules will be fully defined in the table, figure, and data listing specification document.

7. References

1. Ishak K, Baptista A, Bianchi L, Callea F, De Groote J, Gudat F, et al. Histological grading and staging of chronic hepatitis, *J. Hepatol*. 1995;22(6):696-699.
2. Harrison SA, Abdelmalek MF, Caldwell S, Shiffman ML, Diehl AM, Ghalib, R.et al, Simtuzumab is ineffective for patients with bridging fibrosis or compensated cirrhosis caused by nonalcoholic steatohepatitis, *Gastroenterology* 2018;155(4):1140-1153.
3. Hicks KA *et al.* Cardiovascular and Stroke Endpoint Definitions for Clinical Trials. *Circulation* 2017;137**:**961-972, doi:10.1161/CIRCULATIONAHA.117.033502 (2018). https://www.ahajournals.org/doi/full/10.1161/circulationaha.117.033502
4. Hicks KA, et al. standardized definitions for cardiovascular events in clinical trials, 2014 https://www.cdisc.org/system/files/all/standard/Draft Definitions for CDISC August 20%2C 2014.pdf
5. European Medicines Agency, Reflection paper on assessment of cardiovascular safety profile of medicinal products, 2016 https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-assessment-cardiovascular-safety-profile-medicinal-products_en.pdf
6. Zhao Y, Herring AH, Zhou H, Ali MW, Koch GG. A multiple imputation method for sensitivity analyses of time-to-event data with possibly informative censoring. J Biopharm Stat. 2014;24(2):229-53. doi: 10.1080/10543406.2013.860769.

8. Summary of SAP Changes

In addition to minor typographical edits made throughout for clarify, updates from previous SAPs include the following:

- Section 3: updated list of abbreviations and definitions of terms
- Section 4: updated the current versions of the protocol amendment and region-specific protocol amendment that the analyses would be based on
- Section 4.1.1 Overall Design: clarified that the surrogate endpoint for Part 1 analyses are based on histopathological data for the endpoint, and that subjects who continue beyond Month 12 would continue on their same treatment arm throughout the rest of the study. In addition, updated the planned sample size in Part 1 to be approximately 1200
- Section 4.1.1 Part 2: rearranged sentences for clarity and consistency; updated Part 2 randomization information for change in stratification per protocol amendment 4
- Section 4.1.2 Number of Participants: clarified that the Part 1 analysis would commence after completion of the 12-month visit assessments (including biopsy)
- Section 4.2 Study Objectives and Endpoints: Objectives and endpoints were rearranged into global and EU-specific objectives and endpoints, as well as for Part 1 and Part 2 analyses, and endpoints were rephrased for clarity (e.g., summary measures were reworded to be endpoints). [redacted]
- Section 4.3 Schedule of Assessments: updated to current version
- Section 5.1 Statistical Methods Planned in the Protocol and Determination of Sample Size: added sentence about a separate SAP for health economics and outcomes research analyses
- Section 5.1.1 Statistical and Analytical Plans: updated SAS version to be 9.4 or newer

- Section 5.1.1.1.4 Estimands: added estimand language for primary, key secondary, and secondary endpoints
- Section 5.1.1.1.5 Missing Data: separated the data methods for Part 1 analyses from the Part 2 analyses; updated the explanatory variables to include the ELF score at baseline and the weight change from baseline to Month 12
- Section 5.1.1.1.6 Site Pooling: added clarification regarding site as a stratification factor for specified Part 1 analyses
- Section 5.1.1.1.7 Other Common Conventions: clarified what the Part 1 analyses would summarize. Part 1 analyses will include efficacy analyses through Month 12 and available safety data through the Part 1 database lock for the Part 1 participants.
- Section 5.1.1.2.2 Participant Disposition: clarified that the disposition will include thoase who discontinued study drug early or discontinued from the study early for Part 1 or Part 2 of the study. Also, disposition will be presented by total population, treatment group, and by region, but not by study site.
- Section 5.1.1.2.3 Protocol Deviations: corrected language such that significant protocol deviations will be summarized for the ITT population
- Section 5.1.1.2.4 Demographics: removed “Other” for race; any categories for multiple races will be presented under its own multiple race category (e.g., a subject who is both “White” and “Asian” will be presented as “Multiple race: White, Asian”)
- Section 5.1.1.2.5 Baseline Characteristics: clarified fibrosis stage method to be NASH CRN fibrosis stage; clarified baseline definition to be last available assessment prior to study intervention administration; added other disease characteristics like glucagon-like peptide-1 analogue use, sodium-glucose cotransporter-2 inhibitor use, insulin use, categories for liver stiffness, CAP score, ELF score, and key laboratory values; removed others that were not collected in the study (Framingham score, smoking status)
- Section 5.1.1.2.6 Medical History: updated MedDRA version to 22.1 or newer
- Section 5.1.1.2.7 Prior and Concomitant Medications: clarified ATC drug class to be the 4th level, drug code and preferred drug name; clarified that prior and concomitant medications would be presented separately; removed “subsequent medications” parameter, as there will be no additional follow-up on medications after the follow-up visit

- Section 5.1.1.3 Efficacy Analyses: clarification added for the NASH CRN Staging System used for the endpoints; clarified that the Ishak scoring system used in the study is the "modified Ishak scoring system"; added ELF score under fibrosis indices; corrected SF-36 v2 scale names; removed Framingham score, which is not collected in the study; aligned language in SAP to that of the protocol (e.g., protocol used 6 months, SAP updated 180 days to be 6 months)
- Section 5.1.1.3.1 Endpoint Domain 1: Testing hierarchy added for the global analyses as well as region-specific Europe analyses; endpoints, descriptions, and methodologies were updated to reflect previous changes in the SAP amendment
- Section 5.1.1.3.3 Multiple Comparisons Procedure for Primary and Secondary Endpoints: clarified that the alpha control is over both Part 1 and Part 2 of the study with a familywise 0.00125 significance level
- Section 5.1.1.4 Safety Analyses: clarified what the safety endpoints are, along with what the applicable baseline definition was
- Section 5.1.1.4.2 Adverse Events: updated the MedDRA version to be v22.1 or newer; added "grade 2 or above" and "grade 3 or above" as TEAE categories to the overall AE summary table; added treatment-related TESAEs as a category to the overall AE summary table
- Section 5.1.1.4.3.1 Liver Biochemistry: updated section to encompass liver biochemistry elevations, not simply potential Hy's Law or suspected drug-induced liver injuries
- Section 5.1.1.5 Subgroup Analyses: added subgroup analyses for NASH CRN fibrosis stage and T2DM for the key secondary endpoint and other secondary endpoints; clarified geographic region subgroups; removed "diabetic subjects" subset for the subgroup analyses for prior and concomitant medication use; added sodium-glucose cotransporter-2 (SGLT-2) inhibitor use and insulin use subgroups; added weight loss during the study at any time and also weight loss during the study at the liver biopsy assessment as subgroups; removed Framingham score and smoking status subgroup analyses; clarified that the length of biopsy subgroups would be based on both the baseline and Month 12 liver biopsy lengths; updated the alcohol use subgroup categories to be current drinker, former drinker, and non-drinker
- Section 5.1.1.6.1 Part 1 Analyses: clarified that the Part 1 analyses would occur when the approximately 1200 planned participants were followed through the Month 12 visit assessment; clarified that an unblinded team outside of the study team would be performing the unblinded activities through Part 2

- Section 5.2.1 Changes in the Conduct of the Study: approximately 1200 were planned but 1293 were randomized through the screening process during study execution. This would affect the numbers randomized through Part 2, but general conduct of the study is unaffected.

[redacted]

- Section 6.1 Study Treatment Conventions: clarified that the actual fibrosis stage and T2DM status collected via the electronic database would be used for analyses
- Section 6.2.2 Safety: clarified that plasma and serum samples collected beyond Visit 2 would still be analyzable for polymorphisms, genotype status, and pharmacogenetic biobanking
- Section 6.6.1.4 Major Adverse Cardiovascular Events (MACE): added section to describe how potential MACE cases would be identified in the AE dataset; 4-point MACE would be used, and MedDRA SMQs that would be used to identify cases were listed
- Section 6.6.1.5 Type 2 Diabetes Mellitus (T2DM): added section to describe how new onset T2DM cases would be identified in the AE dataset, along with the specific MedDRA preferred terms
- Section 6.6.2.2 Continuous Descriptive Parameters: added the ELF score and PRO-C3
- Section 6.6.3.1 Potentially Clinically Significant Criteria: added PCS criteria for weight change that include abnormal increases or abnormal decreases
- Section 7 References: 3 references added for the MACE criteria
- Section 8 Summary of SAP Changes: added