Document Type:	Final Statistical Analysis Plans
Document Dates:	Statistical Analysis Plan (Surrogate Endpoint Analysis) Version 4.0 – 22 April 2020 Statistical Analysis Plan (Long-Term Endpoint Analysis [Final Analysis]) Version 5.0 – 23 November 2020
Study Title:	A Multicenter, Randomized, Double-Blind, Placebo-Controlled Phase III Study to Evaluate the Efficacy and Safety of Elafibranor in Patients with Nonalcoholic Steatohepatitis (NASH) and fibrosis

Protocol Reference Number: GFT505-315-1

NCT Number: NCT02704403

[redacted]

GENFIT

Protocol No.: GFT505-315-1

A Multicenter, Randomized, Double-Blind, Placebo-Controlled Phase III Study to Evaluate the Efficacy and Safety of Elafibranor in Patients with Nonalcoholic Steatohepatitis (NASH) and fibrosis

[redacted] Study ID: RESOLVE-IT/8325398

STATISTICAL ANALYSIS PLAN

SURROGATE ENDPOINT ANALYSIS

Version: V4.0
Date of Issue: 22-APR-2020

Author: [redacted]

Genfit
Parc Eurasanté
885, Avenue Eugène Avinée
59120 LOOS – France

[redacted]

APPROVALS

The undersigned agree that all required reviews of this document are complete, and approve this Statistical Analysis Plan (SAP) as final. Programming of the tables, figures and listings based upon the specifications within this document can proceed.

[redacted] **Approval:**

[redacted] Date

Printed Name/Title
[redacted]

Genfit Approval:

[redacted] Date

Printed Name/Title
[redacted]

Signature [redacted] Date [redacted]

Printed Name/Title
[redacted]

REVIEWERS

The following reviews of the SAP were conducted:

Name and Title	Role	Version Last Reviewed	Company/ Organization
[redacted]	[redacted]	0.1	[redacted]
[redacted]	[redacted]	0.1	[redacted]
[redacted]	[redacted]	0.1	[redacted]
[redacted]	[redacted]	0.1	[redacted]
[redacted]	[redacted]	0.2	Genfit
[redacted]	[redacted]	0.2	Genfit
[redacted]	[redacted]	0.2	Genfit
[redacted]	[redacted]	0.2	[redacted]
[redacted]	[redacted]	0.3	[redacted]
[redacted]	[redacted]	0.3	Genfit
[redacted]	[redacted]	0.3	Genfit
[redacted]	[redacted]	0.3	[redacted]
[redacted]	[redacted]	0.3	[redacted]
[redacted]	[redacted]	0.4	[redacted]
[redacted]	[redacted]	0.4	Genfit
[redacted]	[redacted]	1.1	Genfit
[redacted]	[redacted]	1.1	Genfit
[redacted]	[redacted]	1.1	[redacted]
[redacted]	[redacted]	1.1	[redacted]
[redacted]	[redacted]	2.0	[redacted]
[redacted]	[redacted]	2.1	[redacted]
[redacted]	[redacted]	2.1	Genfit
[redacted]	[redacted]	2.1	Genfit
[redacted]	[redacted]	2.1	Genfit
[redacted]	[redacted]	2.1	Genfit
[redacted]	[redacted]	2.2	[redacted]
[redacted]	[redacted]	2.2	Genfit
[redacted]	[redacted]	2.2	Genfit
[redacted]	[redacted]	2.3	[redacted]
[redacted]	[redacted]	2.3	Genfit
[redacted]	[redacted]	2.3	Genfit
[redacted]	[redacted]	2.4	[redacted]
[redacted]	[redacted]	2.4	Genfit

[redacted]	[redacted]	2.4	Genfit
[redacted]	[redacted]	2.5	[redacted]
[redacted]	[redacted]	2.5	Genfit
[redacted]	[redacted]	2.5	Genfit
[redacted]	[redacted]	2.6	[redacted]
[redacted]	[redacted]	2.6	Genfit
[redacted]	[redacted]	2.6	Genfit
[redacted]	[redacted]	2.6	Genfit
[redacted]	[redacted]	2.6	Genfit
[redacted]	[redacted]	2.6	[redacted]
[redacted]	[redacted]	2.7	[redacted]
[redacted]	[redacted]	2.7	Genfit
[redacted]	[redacted]	2.7	Genfit
[redacted]	[redacted]	2.7	[redacted]
[redacted]	[redacted]	3.0	[redacted]
[redacted]	[redacted]	4.0	[redacted]
[redacted]	[redacted]	4.0	Genfit
[redacted]	[redacted]	4.0	Genfit
[redacted]	[redacted]	4.0	Genfit
[redacted]	[redacted]	4.0	Genfit

VERSION HISTORY

Version Number	Version Date	Summary and rational of change(s)
2.0	26 Jan 2017	Update with protocol amendment 1 New section added with rules for dealing with missing or incomplete dates
3.0	23 Jan 2020	SAP split into two versions; for the SEA and long-term (final) analysis Clarifications on the SEA timing Additional detail and clarifications provided for analysis set definitions (FITT set, ITT set, ITTF1 set, PPS, FSS, Safety and SSF1) Addition of a new analysis set: Evaluable Efficacy Set Additional section regarding estimand definitions Upgrade of metabolic parameters as Key Secondary Endpoints and set up of a Gatekeeping procedure to control the overall type 1 error rate at 2-sided 0.01 level. Addition of new secondary endpoints Clarifications of time points, visit windows and handling of missing data Update of the statistical method for the primary and sensitivity analyses of the primary endpoint and for the analysis of the binary secondary endpoints defined according to histological parameters. These analyses will be performed using the approach described in Ge et al. (2011) Clarifications on the implementation of the supplementary analysis on the primary endpoint using pattern mixture model Addition of sensitivity analysis on the primary endpoint using CMH test Method of pooling centers clarified (sensitivity analysis including center as fixed effect). Addition of descriptive statistics on demographics and baseline characteristics for the FSS, Safety and SSF1 populations. Addition of descriptive statistics on completion status, baseline characteristics, treatment exposure and treatment emergent adverse events for the subset of patients from the ITT set with missing biopsy results at Week 72 Additional compliance summary up to week 72 for SEA

		Addition of shift tables from baseline to week 72 and follow-up for fibrosis stage, ballooning, inflammation, steatosis, NAS disease activity score and SAF activity score Additional information provided regarding summaries of morphometry data Analyses of change from baseline in NAS disease activity score and Framingham risk scores aligned to have consistent assumptions of variable distribution at SEA and final Analysis of hsCRP and NT-ProBNP updated to be performed on log transformed values Addition of 2 subgroup analyses for primary and key secondary endpoints Addition of exploratory analyses for the ITTF1 population Addition of exploratory analyses for the FITT population Addition of safety analyses for SSF1 population Added raw counts of number of events for summaries of AE data Added clarification regarding which laboratory parameters will be analyzed using repeated measures ANCOVA Added definition and analysis of eGFR New figures for laboratory parameters Updated section 8. "Changes in planned analyses" according to the protocol amendments Editorial changes
4.0	22 Apr 2020	Due to changes in the definition of the long-term primary endpoint, several former components of the endpoint are to be considered as AEs and included in safety analyses Added definition and analysis of AESIs New figure for albuminuria Added clarifications regarding the strategy for eDISH plots Added clarifications regarding the categorization of subjects with serum creatinine increases Added analysis of eGFR decreases Removed the criteria for DILI adjudication that were not consistent with the protocol.

TABLE OF CONTENTS

GLOSSARY OF ABBREVIATIONS

Abbreviation	Term
AE	Adverse Event
AESI	Adverse Event of Special Interest
AKI	Acute Kidney Injury
ANCOVA	Analysis of covariance
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass Index
BP	Blood Pressure
CEC	Clinical Events Committee
CI	Confidence Interval
CMH	Cochran-Mantel-Haenszel
eCRF	Electronic Case Report Form
eGFR	Estimated Glomerular Filtration Rate
CP	Conditional Power
CRA	Clinical Research Associate
CRN	Clinical Research Network
DB	Double-Blind
DILI	Drug Induced Liver Injury
DSMB	Data Safety Monitoring Board
ECG	Electrocardiogram
FAS	Full Analysis Set
FITT	Full Intent-to-treat
FSS	Full Safety Set
F1	Fibrosis Stage 1
F2	Fibrosis Stage 2
F3	Fibrosis Stage 3
GLP	Glucagon-Like Peptide
HCC	Hepatocellular Carcinoma
HR	Heart Rate
ITT	Intent-to-Treat
KDIGO	Kidney Disease Improving Global Outcomes
LOCF	Last observation carried forward
LTTP	Long Term Treatment Period
MAR	Missing at Random
MedDRA	Medical Dictionary for Regulatory Activities
MELD	Model End Stage Liver Disease
NAFLD	Nonalcoholic Fatty Liver Disease
NAS	NAFLD Activity Score
NASH	Nonalcoholic Steatohepatitis
PPS	Per-Protocol Set
PT	Preferred Term
SAE	Serious Adverse Event
SAF	Steatosis-Activity-Fibrosis

SD	Standard Deviation
SE	Standard Error
SEA	Surrogate Endpoint Analysis
SF-36	36-Item Short-Form Health Survey
SI	Standard International
SOC	System Organ Class
SGLT	Sodium/Glucose Cotransporter
TFLs	Tables, Figures and Listings

STATISTICAL ANALYSIS PLAN AMENDMENT 1

Not applicable.

1 SOURCE DOCUMENTS

The Statistical Analysis Plan was written based on the following documentation:

Document	Date	Version
Protocol	15-Jan-2016	1.0
Protocol Amendment	06-Jan-2017	2.0
Protocol Amendment	23-Apr-2017	3.0
Protocol Amendment	20-Apr-2020	5.0
eCRF	22-Feb-2016 07-Apr-2016 05-Apr-2019	1.0 2.0 8.0
DSMB Charter	10-Nov-2016	2.0

This Surrogate Endpoint Analysis Statistical Analysis Plan details objectives, definitions and planned analysis for the surrogate endpoint analysis (SEA). For the same details regarding the interim analysis and final analysis, see the separate Final Analysis Statistical Analysis Plan.

2 PROTOCOL DETAILS

2.1 Study Objectives

2.1.1 Primary objective

The primary objective of the study is to evaluate the efficacy of elafibranor 120 mg versus placebo on resolution of Nonalcoholic Steatohepatitis (NASH) without worsening of fibrosis in patients with fibrosis stage 2 (F2) and fibrosis stage 3 (F3).

Primary objective – surrogate endpoint (at surrogate endpoint analysis)

To evaluate the efficacy of elafibranor 120 mg QD for 72 weeks versus placebo on resolution of NASH without worsening of fibrosis.

- Resolution of NASH is defined as the disappearance of ballooning and disappearance or persistence of minimal lobular inflammation (grade 0 or 1) with an overall pattern of injury not qualifying for steatohepatitis.
- Worsening of fibrosis is evaluated using the NASH Clinical Research Network (CRN) fibrosis staging system and defined as progression of at least one stage.

2.1.2 Secondary objectives

Key secondary objectives (at surrogate endpoint analysis)

To assess histological changes after 72 weeks of treatment on the following endpoint:

- Percentage of patients with improvement of fibrosis of at least 1 stage according to NASH CRN scoring

To assess the clinical benefit after 72 weeks of treatment on the following metabolic endpoints:

- Changes from baseline in Triglycerides, Non-HDL Cholesterol, HDL cholesterol, LDL cholesterol, HbA1c (in diabetic patients), HOMA-IR (in non-diabetic patients)

Other secondary objectives

- To assess histological changes after 72 weeks of treatment on the following endpoints:
  - percentage of patients with improvement of fibrosis of at least 1 stage according to NASH CRN scoring without worsening of NASH
  - percentage of patients with no worsening of Fibrosis and no worsening of NASH
  - percentage of patients with Resolution of NASH and improvement of Fibrosis
  - percentage of patients with at least a 1 point improvement in histological scores (NASH CRN scoring: NAS [sum of steatosis, hepatic ballooning and lobular inflammation], steatosis, hepatic ballooning, lobular inflammation), fibrosis (NAFLD Ishak scoring system), or portal inflammation
  - percentage of patients with improvement of NAS of at least 2 points
  - percentage of patients with improvement of NAS of at least 2 points and with at least 1 point improvement in hepatic ballooning
  - percentage of patients with at least a 1 point improvement in disease activity score (sum of ballooning and lobular inflammation scores) according to NAS scoring and steatosis-activity-fibrosis (SAF) scoring
  - percentage of patients with at least a 1 point improvement in disease activity score (sum of ballooning and lobular inflammation scores) according to NAS scoring and with at least 1 point improvement in hepatic ballooning
  - percentage of patients with at least a 1 point improvement in disease activity score (sum of ballooning and lobular inflammation scores) according to steatosis-activity-fibrosis (SAF) scoring and with at least 1 point improvement in hepatic ballooning
  - percentage of patients with at least 2 points improvement in disease activity score according to NAS scoring and SAF scoring
  - changes in NAS, fibrosis (using NASH CRN and NAFLD Ishak scoring system), steatosis, hepatic ballooning, lobular inflammation, portal inflammation, SAF activity score
  - changes in area of fibrosis (normalized Collagen Proportional area in %) by morphometry
- To assess the following endpoints at Week 72:
  - changes in liver enzymes and liver markers
  - changes in noninvasive markers of fibrosis and steatosis
  - changes in lipid parameters
  - variation in body weight
  - changes in insulin resistance and glucose homeostasis markers
  - changes in inflammatory markers
  - changes in cardiovascular risk profile as assessed by Framingham scores
  - changes in liver stiffness by FibroScan measurement
  - changes in quality of life (36-Item Short-Form Health Survey (SF-36) questionnaire)

Safety secondary objectives

- To assess the tolerability and safety of once a day administration of oral doses of elafibranor 120 mg:
  - adverse events
  - clinical laboratory tests (hematology, chemistry and urinalysis)
  - vital signs
  - electrocardiogram
  - physical examination
  - renal, cardiac, liver, and metabolic parameters
  - other safety markers.

2.1.3 Exploratory objectives

Exploratory objectives

- To constitute a biobank for discovery and validation of biomarkers in NASH.

Exploratory objectives for F1 group

An exploratory objective of the study is to evaluate the efficacy of elafibranor 120 mg versus placebo in a subset of patients with fibrosis stage 1 (F1). The same efficacy and safety endpoints as for the analysis on the F2/F3 cohort will be assessed.

2.2 Overall Study Design

This is a multicenter, randomized, double-blind, placebo-controlled phase III study in patients with NASH.

There are two treatment arms:

- Elafibranor 120 mg
- Placebo

Treatment will be given by oral administration once per day.

Random allocation will be made to the two treatment groups in a 2:1 ratio (elafibranor:placebo) and stratified by the following factors:

- Type 2 diabetes (yes/no)
- Gender (male/female)
- Fibrosis stage (F1/F2/F3)

There will be a 12 week screening period. The first treatment period will last 72 weeks with visits every 12 weeks, and will be followed by a long term treatment period (LTTP) with visits every 24 weeks. The duration of the LTTP will be driven by the number of events and is anticipated to last for approximately 4 years.

The primary comparison of elafibranor versus placebo will be based on a subset of F2 and F3 patients. An additional exploratory comparison will include the subset of F1 patients.

2.3 Sample Size and Power (Surrogate Endpoint Analysis)

2.3.1 Resolution of NASH

The following assumptions were made for the sample size calculation for resolution of NASH:

- α=0.01 two sided
- Randomized patients with no response assessment at Week 72 will be counted as non-responders
- Pooled variance
- Randomization ratio of 2:1 (elafibranor to placebo)
- 8% response rate in control group (placebo)
- 16.5% response rate in active group (elafibranor)

Based on these assumptions, a sample size of 1023 patients in the Intent-to-Treat (ITT) population provides 90% power to show that elafibranor is superior to the placebo with respect to resolution of NASH without worsening of fibrosis.

2.4 Estimand strategy

In line with ICH E9(R1) addendum; four attributes (population, endpoint, intercurrent events and population level summary) have been specified to translate the primary and key secondary efficacy objectives into treatment effect that are to be estimated (estimands).

The intercurrent events as the non-adherence to study treatment, change in diet/exercise, change in dose or initiation of GLP-1 agonists or SGLT-2 inhibitors or statins can be considered as intercurrent events that could alter the interpretation of the biopsy results.
In addition, some others intercurrent events as study treatment discontinuation due to adverse event, refusal of the patients to perform the biopsy can lead to missing biopsy results.

2.4.1 Primary estimand

The primary estimand strategy will be based on the composite strategy imputing non-response for missing data while for the other intercurrent events that could affect the interpretation of the biopsy results the treatment policy strategy will be applied:

- *A. Population:* Randomized patients with fibrosis stage 2 (F2) or fibrosis stage 3 (F3), and NAS $\geq$4
- *B. Endpoint:* Response to treatment (binary variable) indicating a successful response at Week 72 for a patient with resolution of NASH without worsening of fibrosis,
- *C. Intercurrent events:* Intercurrent events leading to missing data will be captured in the Endpoint definition as missing data at Week 72 will be considered as non-responder, whereas the non-missing biopsy results will be considered regardless of the intercurrent events,
- *D. Population-level summary:* Between treatment group difference in response proportions.

In addition, the hypothetical strategy will be also investigated as supplementary estimands and sensitivity analysis (see section 7.7.3) imputing missing reliable biopsy results at Week 72 as if:

- The patients would follow their initial randomised treatment
- The patients from both arms would continue on placebo

2.4.2 Key secondary estimands

- Fibrosis improvement

The same estimand strategy as for the primary estimand will be applied.

- Metabolic parameters improvement

The estimand strategy will be based on the treatment policy strategy:

- *A. Population:* Randomized patients with fibrosis stage 2 (F2) or fibrosis stage 3 (F3), and NAS ≥4
- *B. Endpoint*: Change from baseline to Week 72,
- *C. Intercurrent events:* Regardless of intercurrent events
- *D. Population-level summary:* Between treatment group difference in means.

3 EFFICACY AND SAFETY VARIABLES

3.1 Primary Efficacy Endpoint

Surrogate endpoint - resolution of NASH

The primary surrogate endpoint for this study is resolution of NASH without worsening of fibrosis after 72 weeks of treatment.

Resolution of NASH is defined as the disappearance of ballooning (i.e. grade 0) and disappearance or persistence of minimal lobular inflammation (i.e. grade 0 or 1) with an overall pattern of injury not qualifying for steatohepatitis.

Worsening of fibrosis is evaluated using NASH CRN fibrosis staging system and defined as progression of at least 1 stage.

This surrogate endpoint will be formally assessed at the time of the SEA when at least the first 1023 randomized F2 to F3 patients complete the 72 week treatment period or discontinue early from the study treatment.

3.2 Secondary Efficacy Endpoints

Key Secondary Efficacy Endpoints

- Percentage of patients with improvement of fibrosis of at least 1 stage according to NASH CRN scoring at 72 weeks
- Changes from baseline to Week 72 in metabolic parameters (Triglycerides, Non-HDL Cholesterol, HDL cholesterol, LDL cholesterol, HbA1c (in diabetic patients) and HOMA-IR (in non-diabetic patients))

These key secondary endpoints will be assessed at the time of the SEA when at least the first 1023 randomized F2 to F3 patients complete the 72 week treatment period or discontinue early from study treatment.

Other Secondary Efficacy Endpoints

- Histological changes after 72 weeks of treatment
  - percentage of patients with improvement of fibrosis of at least 1 stage according to NASH CRN scoring without worsening of NASH
  - percentage of patients with no worsening of Fibrosis and no worsening of NASH
  - percentage of patients with Resolution of NASH and improvement of Fibrosis
  - percentage of patients with at least a 1 point improvement in histological scores (NASH CRN scoring: NAS [sum of steatosis, hepatic ballooning and lobular inflammation], steatosis, hepatic ballooning, lobular inflammation), fibrosis (NAFLD Ishak scoring system), or portal inflammation
  - percentage of patients with improvement of NAS of at least 2 points
  - percentage of patients with improvement of NAS of at least 2 points and with at least 1 point improvement in hepatic ballooning
  - percentage of patients with at least a 1 point improvement in disease activity score (sum of ballooning and lobular inflammation scores) according to NAS scoring and steatosis-activity-fibrosis (SAF) scoring
  - percentage of patients with at least a 1 point improvement in disease activity score (sum of ballooning and lobular inflammation scores) according to NAS scoring and with at least 1 point improvement in hepatic ballooning
  - percentage of patients with at least a 1 point improvement in disease activity score (sum of ballooning and lobular inflammation scores) according to steatosis-activity-fibrosis (SAF) scoring and with at least 1 point improvement in hepatic ballooning
  - percentage of patients with at least 2 points improvement in disease activity score according to NAS scoring and SAF scoring
  - changes in NAS, fibrosis (using NASH CRN and NAFLD Ishak scoring system), steatosis, hepatic ballooning, lobular inflammation, portal inflammation, SAF activity score
  - changes in area of fibrosis (normalized Collagen Proportional area in %) by morphometry
- Change from baseline in liver enzymes and liver markers at the end of the 72 weeks treatment period
- Change from baseline in noninvasive markers of fibrosis and steatosis at the end of the 72 weeks treatment period
- Change from baseline in lipid parameters at the end of the 72 week treatment period
- Change from baseline in body weight at the end of the 72 week treatment period

- Change from baseline in insulin resistance and glucose homeostasis markers at the end of the 72 week treatment period
- Change from baseline in inflammatory markers at the end of the 72 week treatment period
- Change from baseline in cardiovascular risk profile as assessed by Framingham scores at the end of the 72 week treatment period
- Change from baseline in liver stiffness by FibroScan measurement at the end of the 72 week treatment period
- Change from baseline in quality of life (SF-36 questionnaire) at the end of the 72 week treatment period

Exploratory Endpoints

- Constitute a biobank for discovery and validation of biomarkers in NASH.

3.3 Safety Variables

The tolerability and safety will be assessed using the following endpoints:

- Adverse events
- Clinical laboratory tests (hematology, chemistry and urinalysis)
- Vital signs
- Electrocardiogram
- Physical examination
- Renal, cardiac, liver, and metabolic parameters
- Other safety markers
- Drug induced liver injury (DILI)

4 PHARMACOKINETIC/PHARMACODYNAMIC VARIABLES

- PK parameters of elafibranor (GFT505) and its metabolite GFT1007 at weeks 12, 24 or 36 for PK population analyses.

The PK analyses will be described in a separate sub study (ancillary study to the protocol). The details for the PK population analysis will be covered in a separate SAP.

5 ANALYSIS POPULATIONS

5.1 Enrolled Set

All patients who signed informed consent.

5.2 Safety Set

All F2 and F3 patients who have taken at least one dose of study treatment will be included in the Safety population. Patients will be analysed according to their actual treatment received.

5.3 Intent-to-treat Set (ITT)

For the purpose of analysis sets aligned with efficacy summaries, a cohort of randomized F2 to F3 patients who complete the Week 72 treatment period or discontinued the study treatment early will be considered.

To characterize this cohort:

- Firstly, the last randomized patient of the cohort will be identified as the F2 or F3 patient who completed the Week 72 treatment period on a pre-defined cut-off date,
- Secondly, all the F2/F3 patients randomized before this patient will be considered as part of the cohort.

Indeed, any F2/F3 patients who prematurely discontinued the study treatment before the pre-defined cut-off date but who were randomized after the last patient of the cohort will not be included in the ITT set. This approach ensures that the ITT set will be consistent with the chronological order and the stratification of the randomization.

Patients will be analyzed according to their randomized treatment.

5.4 Evaluable Efficacy Set (EES)

All F2 and F3 patients in the ITT population who have taken at least one dose of study treatment and have a reliable liver biopsy at both baseline and at the end of the 72 week treatment period. A reliable liver biopsy is one with adequacy status of adequate or marginal. Patients will be analyzed according to their randomized treatment.

5.5 Per Protocol Set (PPS)

All F2 and F3 patients in the EES population who do not have any important protocol deviations leading to exclusion from the PPS.

Protocol deviations are defined as any change, divergence, or departure from the study design or procedures defined in the study protocol. Important protocol deviations are a subset of protocol deviations and may significantly impact the correctness, accuracy, and/or reliability of the study data or that may significantly affect a patient's rights, safety, or well-being. Section 5.5.1 details the deviations.

5.5.1 Important Protocol Deviations Leading to Exclusion from the PPS Analysis

Only those important protocol deviations considered to have a major effect on efficacy will lead to complete exclusion of the patient from the PPS.

Patients will be assessed by comparison of their eCRF data with the criteria below; protocol waivers will not be taken into consideration (e.g. if a patient younger than 18 or older enters the trial on a protocol waiver, the patient would still be excluded from the PPS).

Important protocol deviations will be defined by the Sponsor before database lock, by judging the importance of the influence of the deviations on the primary efficacy endpoint. The following criteria will serve as a basis for the evaluation of the important protocol deviations:

Inclusion and exclusion criteria

Failure to comply with subject inclusion and exclusion criteria that may substantially affect the results of the primary and key secondary efficacy endpoints. These may include:

Inclusion criteria:
Patients must meet all of the following inclusion criteria to be eligible for enrollment in the trial:

1. Histological confirmation of steatohepatitis on a diagnostic liver biopsy by central reading of the slides (biopsy obtained within 6 months prior to Screening or during the Screening Period) with at least 1 in each component of the NAS (steatosis scored 0-3, ballooning degeneration scored 0-2, and lobular inflammation scored 0-3).
2. NAS ≥4.
3. Fibrosis stage of 1 or greater and below 4, according to the NASH CRN fibrosis staging system. For patients with fibrosis stage 1, only patients at high risk of progression will be included meaning with a NAS ≥5 and at least 2 of the following conditions: persistent elevated alanine transferase (ALT; absence of normal value of ALT within the past year), obesity defined by a BMI ≥30, metabolic syndrome (NCEP ATP III definition), type 2 diabetes, or HOMA-IR >6.
4. If a patient is treated with one of the following drugs: vitamin E (>400 IU/day), polyunsaturated fatty acids (>2 g/day), or ursodeoxycholic
Acid, a stable dose from at least 6 months prior to diagnostic liver biopsy is required.
5. For patients with type 2 diabetes, glycemia must be controlled. If glycemia is controlled by anti-diabetic drugs, change in antidiabetic therapy must follow these requirements:
    - no qualitative change 6 months prior to diagnostic liver biopsy up to randomization (i.e. implementation of a new anti-diabetic therapy) for patients
      treated by metformin, gliptins, sulfonylureas, sodium/glucose cotransporter (SGLT) 2 inhibitors, glucagon-like peptide (GLP)-1 agonists, or insulin. Doses changes of these medications are allowed in the 6 months prior to diagnostic liver biopsy, except for GLP-1 agonists, which must remain on stable dose in the 6 months prior to diagnostic liver biopsy.
    - No implementation of GLP-1 agonists or SGLT2 inhibitors up to 72 weeks of treatment (V7).
      Initiation of any other antidiabetic drugs is allowed after Randomization based on treating physician's judgment, except for glitazones which are prohibited 6 months prior to diagnostic liver biopsy until the end of treatment.

Exclusion criteria:
Patients who meet any of the following criteria will be excluded from entering the study:

1. Known chronic heart failure (Grade I to IV of New York Heart Association classification).
2. History of efficient bariatric surgery within 5 years prior to Screening, or planned bariatric surgery in the course of the study.
3. Uncontrolled hypertension during the Screening Period despite optimal antihypertensive therapy.
4. Type 1 diabetes patients.

5. Patients with HbA1c >9.0%. If abnormal at the first Screening Visit, the HbA1c measurement can be repeated at the latest 2 weeks prior to Randomization. A repeated abnormal HbA1c (HbA1c >9.0%) leads to exclusion.
6. Patients receiving thiazoledinediones (glitazones [pioglitazone, rosiglitazone]), unless the drug was discontinued at least 6 months before the diagnostic liver biopsy.
7. Patients with a history of clinically significant acute cardiac event within 6 months prior to Screening such as: stroke, transient ischemic attack, or coronary heart disease (angina pectoris, myocardial infarction, revascularization procedures).
8. Weight loss of more than 5% within 6 months prior to Randomization.
9. Compensated and decompensated cirrhosis (clinical and/or histological evidence of cirrhosis). Notably, NASH patients with fibrosis stage = 4 according to the NASH CRN fibrosis staging system are excluded.
10. Current or recent history (<5 years) of significant alcohol consumption. For men, significant consumption is typically defined as higher than 30 g pure alcohol per day. For women, it is typically defined as higher than 20 g pure alcohol per day.
11. Pregnant or lactating females or females planning to become pregnant during the study period.
12. Other well documented causes of chronic liver disease according to standard diagnostic procedures including, but not restricted to:
    - Positive hepatitis B surface antigen (HBsAg)
    - Positive HCV RNA, (tested for in case of known cured HCV infection, or positive HCV Ab at Screening)
    - Suspicion of drug-induced liver disease
    - Alcoholic liver disease
    - Autoimmune hepatitis
    - Wilson's disease
    - Primary biliary cirrhosis, primary sclerosing cholangitis
    - Genetic homozygous hemochromatosis
    - Known or suspected HCC
    - History or planned liver transplant, or current MELD score >12.
13. Known hypersensitivity to the investigation product or any of its formulation excipients.
14. Evidence of any other unstable or, untreated clinically significant neoplastic disease.

Compliance

- Treatment compliance less than 80% or greater than 120%. For the SEA this applies for data over the 72 week double blind treatment period.

Prohibited procedures

Prohibited procedures include:
- Bariatric surgery up until the week 72 liver biopsy.

Prohibited medications

Prohibited medications include:

- Any medication that can induce steatosis/steatohepatitis including, but not restricted to: corticosteroids (parenteral & oral chronic administration only), amiodarone (Cordarone), tamoxifen (Nolvadex), and methotrexate (Rheumatrex, Trexall) within 30 days prior to Screening and up to the end of treatment.
- Any medication that could interfere with study medication absorption, distribution, metabolism, or excretion or could lead to induction or inhibition of microsomal enzymes, e.g. indomethacin, from randomization and up to the end of study treatment visit.
- GLP-1 agonist dose stability required from at least 6 months prior to the inclusion liver biopsy up to end of treatment visit.
- Initiation of GLP-1 agonist and SGLT2 inhibitors is prohibited during the study up to 72 weeks of treatment.
- Implementation of new anti-diabetic therapy (insulin, sulfamides, metformin, gliptins, SGLT2-inhibitors) from at least 6 months prior to the inclusion of the liver biopsy.
- Lipid lowering therapy (Statins, Ezetimibe) dose stability is required from at least 2 months prior to screening.
- Fibrates are prohibited 2 months before Randomization up to end of treatment.
- Vitamin E >400 IU/day dose stability is required from at least 6 months prior to diagnostic liver biopsy.
- Polyunsaturated fatty acids > 2 g/day dose stability is required from at least 6 months prior to diagnostic liver biopsy.
- Ursodeoxycholic acid dose stability is required from at least 6 months prior to diagnostic liver biopsy.

Only medications which are administered up until the week 72 liver biopsy will be considered for potential impact upon the PPS for the SEA.

Minimum exposure to treatment

- Failure to receive at least 58 weeks of study treatment.

Availability of primary efficacy data

- No liver biopsy test at week 72.

Errors in treatment allocation

- Received incorrect treatment compared with randomized treatment. Only errors in treatment allocation up until the week 72 liver biopsy will be considered for exclusion from the PPS at the time of the SEA.

Unblinding

- Unblinding of treatment. Only unblinding incidents up until the week 72 liver biopsy will be considered for exclusion from the PPS at the time of the SEA.

Important protocol deviations leading to exclusion from the PPS will be determined programmatically from the data and recorded by clinical research associates (CRA) at site. Those criteria which require clinical or medical monitoring interpretation will be reviewed prior to database lock.

All important protocol deviations leading to exclusion from the PPS occurring during the study will be reviewed and approved by Genfit on a case-by-case basis prior to database lock and unblinding. Should additional important protocol deviations leading to exclusion from the PPS, not anticipated at the time of preparing this SAP, be identified during the study and prior to unblinding they will be documented in a PPS Identification Plan and included in all relevant protocol deviation reviews and approvals.

Patients in the PPS will be analyzed according to their randomized treatment.

5.6 Other Populations

5.6.1 Full Safety Set (FSS)

All patients who have taken at least one dose of study treatment. Patients will be analyzed according to their actual treatment.

5.6.2 Full Intent-to-Treat Set (FITT)

All patients in the ITT Set, and all F1 patients who were randomized on or prior to the last randomized date of the patients in the ITT Set.
Patients will be analyzed according to their randomized treatment.

5.6.3 Intent-to-Treat F1 Cohort Set (ITTF1)

All F1 patients in the FITT. Patients will be analyzed according to their randomized treatment.

5.6.4 Safety F1 Cohort Set (SSF1)

All F1 in the FSS. Patients will be analyzed according to their actual treatment.

6 DATA HANDLING

6.1 Time points and Visit Windows

The following rules will be used for the by-visit analyses of the liver enzymes and liver markers, non-invasive markers of fibrosis and steatosis, lipid parameters, weight, insulin resistance and glucose homeostasis markers, inflammatory markers, SF-36 questionnaire, clinical laboratory tests, vital signs, ECG, liver stiffness by FibroScan measurement and other safety markers.

For all populations, assessments will be assigned to visits for non-categorical summaries as follows:

- Only assessments recorded with a nominal visit number or recorded as an unscheduled visit will be considered for assignment to visits (i.e. results recorded on the end of treatment and end of study CRF forms will not be used).
- For baseline* assessments, the latest assessment with non-missing result on or before study day 1 will be used. Note that this may include visits also assigned to screening for by-visit summaries.
- For post-baseline assessments, only those up to and including 30 days post last dose of study treatment will be considered for assignment to visits.
- Assessments with missing data and assessments marked 'Not Done' will be as providing a missing response and are not permitted to be assigned to a visit window.
- If the initial nominal visit, (defined hereafter as the earliest assessment for a given visit to be recorded as the nominal visit number via the CRF), has a non-missing response, then that will be chosen for analysis.
- If the initial nominal visit has a missing response, then the earliest of all subsequent assessments with a non-missing response which have either:
    - been recorded as the same nominal visit, or
    - been recorded as an unscheduled visit which is assigned to the given visit as per Table 1

  will be chosen for the analysis.
- If a given visit has no initial nominal visit record, then the earliest of all unscheduled visits which are assigned to the given visit as per Table 1 will be chosen for the analysis.

For all populations, assessments will be assigned to visits for safety categorical summaries (e.g. clinical laboratory data) as follows:

- Only assessments recorded with a nominal visit number or recoded as an unscheduled visit will be considered for assignment to visits.
- Assessments with missing data and assessments marked 'Not Done' will be as providing a missing response and are not permitted to be assigned to a visit window.
- The worse value (e.g. out of reference range preferred to within range; abnormal preferred to normal) will be used in each window

End of the 72 week treatment period visit will be defined as the last post-baseline visit during the 72 week double-blind treatment period.

Day 1 will be defined as the day of first dose of double-blind study treatment. For patients randomized and not treated, Day 1 will be defined as the day of randomization.
Relative days after Day 1 are calculated as (assessment date – Day 1 date) + 1. Relative days prior to Day 1 are calculated as (assessment date – Day 1 date). The day prior to Day 1 is Day -1.

Table 1 Definition of visit windows

Visit	Target Day of Visit[a]	Acceptable visit window
1 Screening	Day –X[b]	Up to Day -1
2 Baseline	Day 1	*See baseline definition above
3 Week 12	Day 84	Days 2 to 125
4 Week 24	Day 168	Days 126 to 209
5 Week 36	Day 252	Days 210 to 293
6 Week 48	Day 336	Days 294 to 377
7 Week 60	Day 420	Days 378 to 461
8 Week 72	Day 504	Days 462 to 587
9 Week 96	Day 672	Days 588 to 755
10 Week 120	Day 840	Days 756 to 923
11 Week 144	Day 1008	Days 924 to 1091
12 Week 168	Day 1176	Days 1092 to 1259
13 Week 192	Day 1344	Days 1260 to 1427
14 Week 216	Day 1512	Days 1428 to 1595
15 Week 240	Day 1680	Days 1596 to 1763
16 Week 264	Day 1848	Days 1764 to 1931
17 Week 288	Day 2016	Days 1932 to 2099
18 Week 312	Day 2184	Days 2100 to 2267
End of the 72 week treatment period	N/A	Last post-baseline visit during the 72 week double-blind treatment period

[a] relative to the date of first dose of double-blind study treatment (Day 1)
[b] screening visit has different target days depending upon the assessment performed

All data will be listed. However, only data slotted into the protocol defined visits will be reported in table summaries. For example, ECG data will only be summarized at Baseline, Week 36, Week 72, Week 120 etc.

All other analysis will use the nominal study visit as defined in the Study Schedule and eCRF. For example, liver biopsy results will be summarized at baseline and Week 72.

6.2 Handling of Dropouts or Missing Data

6.2.1 Handling of Missing Data

Patients with missing data for resolution of NASH without worsening of fibrosis will be treated as a non-responder for the main analysis. In general, missing outcomes for categorical efficacy endpoints will be treated as non-responders/ treatment failures. For continuous endpoints, inference will be done using mixed-effects models repeated-measures, which assumes that missing outcomes are missing at random (MAR).

Supplementary analyses using multiple imputations and a pattern mixture model will be performed on the primary efficacy endpoint at the time of the SEA.

Firstly, missing values for resolution of NASH without worsening of fibrosis will be imputed multiple times to account for the uncertainty about the true values to impute. Missing values will be imputed using the monotone regression method with one hundred imputed datasets. See section 7.7.3 for further details.

Secondly, to assess the sensitivity of the results to non-ignorable missing data, missing values for resolution of NASH without worsening of fibrosis will be imputed using a pattern mixture model. This analysis uses the placebo group as the basis for imputation of any missing values. See section 7.7.3 for further details.

Patients with missing data for improvements in liver biopsy results (e.g. improvement of fibrosis according to NASH CRN of at least one grade) will be treated as non-responders.

All other data (e.g. non-binary [i.e. other than responder / non-responder] biopsy assessment data recorded at week 72, study population/baseline characteristics and adverse events) will be analysed using a complete case approach, whereby the missing data will be ignored and the analysis will be performed using only the data which was actually collected.

6.2.2 Handling of Missing or Incomplete Dates

Incomplete dates (partial or missing dates) will be presented in the data listings as provided on the electronic Case Report Form (eCRF). However, for use in calculations (e.g. to calculate the duration of an AE or medication use) dates will be estimated as follows:

For partial start dates:

If the year is unknown, then:

- The date will not be imputed and will be assigned a missing value.

If the month is unknown, then:

- If the year matches the year of the first dose date, then impute the month and day of the first dose date.
- Otherwise assign “January”.

If the day is unknown, then:

- If the month and year match the month and year of the first dose date, then impute the day of the first dose date.
- Otherwise assign “01”.

For partial end dates:

- If the year is unknown, then the date will not be imputed and will be assigned a missing value.
- If the month is unknown, then assign “December”.
- If the day is unknown, then assign the last day of the month.

If the above rules for end-dates result in illogical date with regards to the dates the subject was in the study, then the end date will be replaced with the subject's date of completion/discontinuation.

7 STATISTICAL METHODS

7.1 Surrogate Endpoint Analysis

The SEA will take place when at least the first 1023 randomized F2 and F3 patients have completed the 72 week treatment period or discontinued early from the study treatment.

At this time, a snapshot of the database will be cleaned and locked for analysis and potential Subpart H or conditional approval submission. This analysis will be performed by an unblinded team separate from the study team. The study team will remain blinded until the final analysis at the end of the follow-up period.

All the efficacy data collected over the Week 72 treatment period will be considered for the efficacy analyses. In addition, any Week 72 biopsy performed until 42 days (6 weeks) after the pre-defined cut-off date used to determine the ITT set (see section 5.3) will be included in the efficacy data for the SEA.

All safety data collected up until the date of the last visit over the Week 72 treatment period from the ITT set will be summarized in safety outputs.

7.2 General Principles

All data processing, summarization and analyses will be performed using the Hosted SAS Environment / Version 9.1.3 (or later) of the SAS® statistical software package.

Unless specified otherwise, data will be displayed using the following treatment group labels, in the order presented:

- Elafibranor
- Placebo

All data collected will be presented in listings by treatment group, country, patient, assessment and visit (where applicable), unless otherwise specified.

Data will be presented in summary tables by treatment group, assessment and visit (where applicable).

The category "Missing" will be presented if the number missing is greater than zero for at least one treatment group.

Descriptive summary statistics for continuous variables will include the number of patients (n), mean, standard deviation (SD), median, minimum and maximum.

In order to calculate summary statistics for log-transformed continuous variables, results will be log-transformed at the patient level and summary statistics will be calculated after back-transformation to

provide the geometric mean and associated standard error (SE). Descriptive summary statistics for log-transformed continuous variables will include the n, geometric mean, SE, median, minimum and maximum.

Descriptive summary statistics for categorical variables will include frequency counts and percentages [n (%)]. Unless stated otherwise in the table shells, the denominator for percentage calculations will be the number of patients in the analysis population.

Change from baseline will be calculated where both baseline and corresponding post-baseline assessments are present.

Dates will be displayed as DDMMMYYYY.

Significance tests for the surrogate endpoint (resolution of NASH without worsening of fibrosis) will be tested at the SEA and will use a two-sided 1% significance level. If this test is significant, the key secondary efficacy endpoints will be tested at a two-sided 1% significance level according to the gatekeeping procedure defined in section 7.7.

All other significance tests will be two-sided and use a 5% significance level for main effects and 10% significance level for interaction terms.

7.3 Patient Disposition and Data Sets Analysed

Patient disposition will be listed and summarized by treatment group and overall.

The number and percentage of patients in the following categories will be summarized for patients in the enrolled population:

- screened
- randomized
- non-randomized (screen failures)
- randomized and not treated
- treated

The number and percentage of patients in each of the analysis populations (Enrolled, ITT, EES, PPS, FITT, ITTF1, Safety, FSS and SSF1) will be summarized by treatment group and overall for patients in the enrolled population.

The number and percentage of patients who complete the study and who discontinue the study early, including a breakdown of the primary reasons for discontinuation, will be presented for the ITT, FITT, ITTF1, Safety, FSS and SSF1 populations. In addition, the number and percentage of patients who complete the treatment and who discontinue the treatment early, including a breakdown of the primary reasons for discontinuation of treatment early will be presented.

For the ITT, FITT and ITTF1 populations, as well as in the subset of patients from the ITT set with missing biopsy results at Week 72, the completion status of patients (early study discontinuation and early study

treatment discontinuation) with the primary reasons for discontinuation will also be displayed over the 72 week treatment period.

A summary of patient enrollment by country and site will also be provided by treatment group and overall for the ITT, FITT, ITTF1, Safety, FSS and SSF1 populations.

A summary of the reasons for screen failure will be produced for patients in the enrolled population. No other information for screen failures will be presented.

The reason for exclusion from the EES will also be summarized by treatment group for the ITT population.

A summary of patients who attended each visit will be produced for the ITT, FITT, ITTF1, Safety, FSS and SSF1 populations.

7.4 Protocol Deviations

All important protocol deviations will be listed for the FITT population.

All important protocol deviations (see Section 5.5.1) will be summarized by treatment group for the ITT, FITT, ITTF1, Safety, FSS and SSF1 populations. All important protocol deviations leading to exclusion from the PPS population will be summarized by treatment group for the EES population.

The important protocol deviations will be identified before data are unblinded.

7.5 Demographics and Other Baseline Characteristics

Demographic and baseline characteristics will be listed and summarized by treatment group and overall for patients in the ITT, FITT, ITTF1, Safety, FSS and SSF1 populations.

These characteristics will also be described in the subset of patients from the ITT set with missing biopsy results at Week 72.

Standard descriptive statistics will be presented for the continuous variables of:

- age (years);
- weight (kg);
- height (cm);
- body mass index (kg/m^2) [calculated as ($weight/height^2$) where weight is in kg and height is in m];
- waist circumference (cm);
- NAS (calculated as the sum of steatosis, hepatic ballooning and lobular inflammation);
- NAFLD fibrosis score;
- Model end stage liver disease score (MELD) score;

The total counts and percentages of patients will be presented for the categorical variables of:

- age group (years) (grouped as <60, and ≥60);

- gender (male, female);
- race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Pacific Islander, White);
- ethnicity (Hispanic or Latino, Not Hispanic or Latino);
- type 2 diabetes (yes, no);
- fibrosis stage (F1, F2, F3);
- child bearing potential (yes, no, not applicable);
- NAS severity [grouped as moderate (4-5), and severe (≥6)];
- MELD score (<15, ≥15);

No formal tests of statistical significance will be performed on the demographic and baseline data.

Other baseline measurements, such as vital signs, will be summarized by treatment group with the post-baseline measurements.

7.5.1 Medical History

Medical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) [Version 19.0 (or a later version if updated during the study)]. All medical history will be listed, and the number and percentage of patients with any medical history will be summarized for patients in the ITT, FITT, ITTF1, Safety, FSS and SSF1 populations by system organ class (SOC) and preferred term (PT) for each treatment group and overall.

7.5.2 Prior and Concomitant Medications

Medications received prior to or concomitantly with study treatment will be coded by [redacted] using the WHO Drug Dictionary [Version B2-BDE March 2016 (or a later version if updated during the study)], Anatomical Therapeutic Chemical (ATC) Classification codes.

Prior medications and concomitant medications are defined as follows:

Prior medications are those taken with a stop date prior to the first dose date of study treatment.

Concomitant medications are those with a start date on or after the first dose date of study treatment, or those with a start date before the first dose date of study treatment and a stop date on or after the first dose date of study treatment.

If a medication cannot be classified as "prior" or "concomitant" after applying imputation rules for missing/incomplete dates, it will be classified as concomitant.

Prior medications and concomitant medications will be listed together and summarized separately for patients in the ITT, FITT, ITTF1, Safety, FSS and SSF1 populations.

The number and percentage of patients using each medication will be displayed together with the number and percentage of patients using at least one medication within each therapeutic class (ATC-Level 2), chemical subgroup (ATC-Level 4), and generic term.

7.5.3 Substance Use History

Substance use history will be listed and summarized by treatment group and overall for patients in the ITT, FITT, ITTF1, FSS, Safety and SSF1 populations.

The total counts and percentages of patients will be presented for the categorical variables of:

- Does the patient smoke or have a history of smoking? (Yes, No)
- Smoking status (Current, Former)
- If current or former smoker, type of substance (Cigarettes, Pipes, Cigars, Other)
- Does the patient consume alcohol? (Yes, No)
- If consume alcohol, frequency of alcohol consumption (< 1-7 drink units per week, 8 drink units per week, 9-14 drink units per week, 15 or more drink units per week)
- Does the patient consume caffeinated coffee? (Yes, No)
- Frequency of coffee consumption (< 1-7 cups per week, 8 cups per week, 9-14 cups per week, 15 or more cups per week)

In addition, the number of cigarettes / pipes / cigars smoked per day will be summarised for current smokers.

7.5.4 Safety, Diet and Lifestyle Factors

Safety Diet and Lifestyle Factors will be listed and summarized by visit, treatment group and overall for patients in the ITT, FITT, ITTF1, FSS, Safety and SSF1 populations.

The total counts and percentages of patients will be presented for the categorical variables of:

- Significant changes in diet since last visit? (Yes, No)
- Significant change in physical activity since last visit? (Yes, increased, Yes, decreased, No)
- Does the patient consume alcohol? (Yes, No)
- Type of alcohol consumption, (Beer, Wine, Distilled spirits or liquor, Other)
- Frequency of alcohol consumption (< 1-7 drink units per week, 8 drink units per week, 9-14 drink units per week, 15 or more drink units per week, Other)
- Significant change in dietary supplement intake? (Yes, No)
- Does the patient consume caffeinated coffee? (Yes, No)
- Frequency of coffee consumption (< 1-7 cups per week, 8 cups per week, 9-14 cups per week, 15 or more cups per week)
- Current smoking status (Smoker, Non-smoker)
- Occurrence of diabetes? (Yes, No)
- Currently on medication for high blood pressure (Yes, No)
- Patient still compliant with study drug intake (Yes, No, Not applicable)

7.6 Measurements of Treatment Compliance

Percentage compliance will be calculated as:

100 * actual tablets taken/expected tablets taken

Where actual tablets taken is defined as the sum of the tablets taken and tablets expected to be taken is defined as (date of last dose – date of first dose) + 1.

A further measure of percentage compliance will be defined over the 72 week double blind treatment period, where:

- actual tablets taken is defined as the sum of the tablets taken up to and including the week 72 assessment, and
- expected tablets taken is defined as (date of last dose up to and including week 72 assessment – date of first dose) + 1.

Percentage compliance and percentage compliance up until week 72 will be separately summarized descriptively by treatment group for patients in the ITT, FITT, ITTF1, Safety, FSS and SSF1 populations.

The number and percentage of compliant patients (both up until week 72 and overall) will be presented for the ITT, FITT, ITTF1, Safety, FSS and SSF1 populations, where compliant is defined as percentage compliance between 80.0% and 120.0% inclusive. The following percentage compliance categories will be presented:

- <50.0%
- ≥50.0% to <80.0%
- ≥80.0% to ≤120.0%
- >120.0%

7.7 Efficacy

The primary surrogate endpoint is resolution of NASH without worsening of fibrosis after 72 weeks of treatment. This endpoint will be analysed at the time of the SEA, when at least the first 1023 randomized F2 and F3 patients have completed the 72 week treatment period or discontinued from the study treatment.

The main objective of the study is to show superiority of elafibranor compared to placebo. The overall type 1 error for the primary endpoints (primary surrogate endpoint at Week 72 and primary long term endpoint) is two-sided alpha=0.05. To control the overall type 1 error for the two primary endpoints, the overall alpha risk will be split 20%:80%, with two sided alpha=0.01 for the primary surrogate endpoint.

The key secondary efficacy endpoints will be tested only if the primary surrogate endpoint is statistically significant. A gatekeeping procedure will be constructed to control the overall Type I error rate for testing the key secondary efficacy endpoints at an overall two-sided alpha level of 0.01.
The key secondary endpoints will be defined as follows:

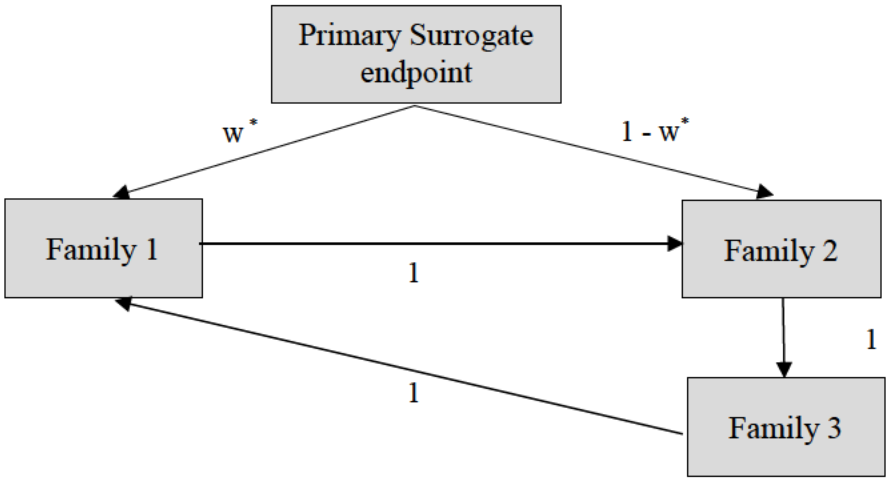
- Endpoint 1: Percentage of patients with improvement of fibrosis of at least 1 stage according to NASH CRN scoring.
- Endpoints 2: Change in metabolic parameters
  - Endpoint 2.1: Triglycerides.

- Endpoint 2.2: Non-HDL cholesterol.
- Endpoint 2.3: HDL cholesterol.
- Endpoint 2.4: LDL cholesterol.
- Endpoint 2.5: HbA1c (diabetic subpopulation).
- Endpoint 2.6: HOMA IR (non-diabetic subpopulation).

The resulting seven null hypotheses of no treatment effect will be labeled as follows:

- Hypothesis H_1: No treatment effect with respect to Endpoint 1.
- Hypothesis $H_{2.1}$: No treatment effect with respect to Endpoint 2.1.
- …
- Hypothesis $H_{2.6}$: No treatment effect with respect to Endpoint 2.6.

The gatekeeping procedure will be set up using the general method for building multi-stage parallel gatekeeping procedures in multiplicity problems with several families of null hypotheses (Dmitrienko and Tamhane, 2011, 2013).
The null hypotheses will be grouped into three families as shown in Figure 1 (Families 1 through 3). Family 1 will include the hypothesis H_1 , Family 2 the hypotheses $H_{2.1}$, $H_{2.2}$, $H_{2.3}$ and Family 3 the hypotheses $H_{2.4}$, $H_{2.5}$, $H_{2.6}$.
The families will be tested as shown in the following Figure:



* *The weight w will be defined equal to 0.2*

The gatekeeping procedure will rely on the following stepwise algorithm:

- *Step 1*. Test the hypothesis H_1 in Family 1 at a two-sided $\alpha_1 = \alpha w$, where $\alpha = 0.01$. Proceed to Step 2.
- *Step 2*. Test the hypotheses included in Family 2 using the truncated Holm test with the truncation parameter γ^{**}. The truncated Holm test will be carried out at a two-sided $\alpha_2 = \alpha$ if the hypothesis H_1 is rejected in Step 1 or at a two-sided $\alpha_2 = \alpha(1 - w)$ if the hypothesis H_1 is not rejected in Step 1. Proceed to Step 3 if at least one hypothesis is rejected in Family 2.

- *Step 3*. Test the hypotheses included in Family 3 using the truncated Holm test with the truncation parameter γ^{**}. The truncated Holm test will be carried out at a two-sided $\alpha_3 = \alpha_2$ if all hypotheses are rejected in Family 2 in Step 2 or at a two-sided $\alpha_3 = \alpha_2(1-\gamma)r_2/m_2$ if some hypotheses are not rejected in Family 2 in Step 2, where r_2 is the number of rejected hypotheses in Family 2 and m_2 is the total number of hypotheses in Family 2. Proceed to Step 4 if at least one hypothesis is rejected in Family 3.
- *Step 4*. Return to Family 1 if the hypothesis H_1 is not rejected in Step 1 and at least one hypothesis is rejected in Family 3. Test the hypothesis H_1 at a two-sided $\alpha_1 = \alpha w + \alpha_3$ if all hypotheses are rejected in Family 3 in Step 3 or at a two-sided $\alpha_1 = \alpha w + \alpha_3(1-\gamma)r_3/m_3$ if some hypotheses are not rejected in Family 3 in Step 3, where r_3 is the number of rejected hypotheses in Family 3 and m_3 is the total number of hypotheses in Family 3.

** *The truncated parameter γ will be defined equal to 0.8*

The resulting gatekeeping procedure guarantees overall Type I error rate control across the pre-defined key secondary hypotheses at a two-sided $\alpha = 0.01$

Statistical testing for other secondary efficacy endpoints will be exploratory. Although treatment differences, 95% CIs and p-values will be presented, no formal confirmatory testing will be performed.

7.7.1 Primary Efficacy Analysis

The primary surrogate endpoint is resolution of NASH without worsening of fibrosis after 72 weeks of treatment.

Resolution of NASH is defined as the disappearance of ballooning and the disappearance or persistence of minimal lobular inflammation (grade 0 or 1) with an overall pattern of injury not qualifying for steatohepatitis.

Worsening of fibrosis is evaluated using the NASH Clinical Research Network (CRN) fibrosis staging system and is defined as progression of at least one stage.

Information about resolution of NASH and worsening of fibrosis will be documented as part of the liver biopsy results.

The main confirmatory analysis will take place at the time of the SEA, when at least the first 1023 F2 and F3 randomized patients have completed the 72 week treatment period or discontinued from the study treatment.

The main analysis will be performed on the ITT population. Supportive analysis will be performed on the EES and PPS populations.

The null hypothesis is that there will be no difference in response rates between the elafibranor and placebo treatment groups. The alternative hypothesis is that there will be a difference in the response rates between the elafibranor and placebo treatment groups. The null hypothesis will be tested at the two-sided 0.01 alpha level.

The number and percentage of patients with resolution of NASH without worsening of fibrosis will be summarized by treatment group for patients in the ITT, EES and PPS populations.

Patients with missing data for resolution of NASH without worsening of fibrosis will be treated as non-responders for the main analysis.

The estimand (the population-average difference in the response rates based on resolution of NASH without fibrosis) will be estimated using the method described in Ge et al. (2011), Appendix A. Specifically, the logistic model with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3) and baseline NAS (categorical) will be fitted.

$$logit\big(P(y_i = 1)\big) = \beta_0 + \beta_1 t_i + x_i^T \beta_2$$

Here $x_i, i = 1, \ldots, N$ is a k-dimensional vector of covariates, t=0/1 is the dummy variable coding the experimental treatment versus control, and $\hat{\hat{\beta}}$ is k+2 dimensional vector of coefficients including intercept, β_0, coefficient for the treatment variable $\beta_1$and a k dimensional vector β_2. For each patient i, two potential outcomes will be computed using the estimated logistic regression coefficients: the probability of response if the patient was assigned to the experimental treatment $p_{i,1} = P_{\hat{\beta}}(Y = 1|x = x_i, t = 1)$ and the probability of response if the patient was assigned to the control, $p_{i,0} = P_{\hat{\beta}}(Y = 1|x = x_i, t = 0)$. The population average estimate of the treatment effect is computed as $d = N^{-1}\sum_{i=1}^{N} p_{i1} - N^{-1}\sum_{i=1}^{N} p_{i0}$. The variance of d can be estimated using the delta method. Using the notation proposed in Ge et al. (2011), let V be the (k+2) by (k+2) covariance matrix of the estimated coefficients $\widehat{\beta}$. Let N-dimensional vector P_1 comprise elements $p_{i,1}$ and P_0 comprise elements $p_{i,0}$. Introduce the derivative vectors A_1 and A_0 with elements, $a_{i0} = p_{i0}(1 - p_{i0})$ and $a_{i1} = p_{i1}(1 - p_{i1})$. Also introduce N by (k+2) the modified design matrices $\tilde{X}_0$ and $\tilde{X}_1$ with columns representing a vector of 1's, vector of treatment indicators and k baseline covariates. For $\tilde{X}_0$ set all values for treatment indicators to 0 and for $\tilde{X}_1$ to 1. Let Then $d_0 = A_0^T\tilde{X}_0/N$ and $d_1 = A_1^T\tilde{X}_1/N$. Then, the variance of d is approximated by $var(d) = d_0 \mathrm{V}\, d_0^T + d_1 \mathrm{V}\, d_1^T - 2d_1 \mathrm{V}\, d_0^T$.

The analysis will be performed using the SAS code included in Ge et al. (2011), Appendix C.

Thus, the estimate of difference (elafibranor/placebo) in rate of resolution of NASH without worsening of fibrosis, its 99% CI and corresponding p-value will be presented. The p-value will be computed using the normal approximation method. In addition, the 95% CI will be presented. The rate of resolution of NASH without worsening of fibrosis will also be presented for each treatment group with unadjusted 95% CIs using the Wilson score method.

The Hosmer-Lemeshow test will be used to assess the goodness of fit of the fitted logistic model.

Further supplementary and sensitivity analysis are described in section 7.7.3.

7.7.2 Secondary Efficacy Analysis

7.7.2.1 Key Secondary Efficacy Analysis

- Improvement of fibrosis according to NASH CRN scoring after 72 weeks of treatment

This endpoint will be analysed at the time of the SEA, when at least the first 1023 randomized F2 and F3 patients have completed the 72 week treatment period or discontinued from the study treatment.

Improvement of fibrosis according to NASH CRN scoring is defined as an improvement in NASH CRN scoring of at least one stage from baseline, where an improvement is defined as a reduction in stage (i.e. a decrease). For example, if a subject had Stage 3 fibrosis according to NASH CRN scoring at the liver biopsy taken at baseline and Stage 2 fibrosis according to NASH CRN scoring at the liver biopsy taken at the end of the 72 week treatment period this will be defined as an improvement in fibrosis of at least one stage according to NASH CRN scoring. Improvements require a change within the numerical component of the fibrosis stage 1 (e.g.) a change from 1c to 1b, or a change from 1b to 1a, does not constitute an improvement for fibrosis wherever this term is used.

The number and percentage of patients with improvement of fibrosis according to NASH CRN scoring at the end of the 72 week treatment period will be summarized by treatment group for patients in the ITT, EES and PPS populations.

Patients with missing data for improvement of fibrosis according to NASH CRN scoring will be treated as non-responders for the main analysis of this endpoint.

The number of patients with an improvement of fibrosis according to NASH CRN scoring at the end of the 72 week treatment period will be analysed using the same method as for the primary analysis of the primary endpoint (see section 7.7.1).

The estimate of difference (elafibranor/placebo) in rate of improvement of fibrosis according to NASH CRN scoring, its 99% CI, corresponding raw p-value and multiplicity adjusted p-value computed according to the pre-defined gatekeeping procedure will be presented. In addition, the 95% CI will also be presented. The rate of improvement of fibrosis according to NASH CRN scoring will also be presented for each treatment group with unadjusted 95% CIs using the Wilson score method.

The main analysis of this key secondary endpoint will be performed on the ITT population. Supportive analyses will be performed on the EES and PPS populations.

- Improvement of metabolic parameters after 72 weeks of treatment

The change from baseline in metabolic parameter values at each time-point will be analysed using a repeated measures ANCOVA model with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3), baseline endpoint specific parameter value, time-point, and treatment by time-point interaction. Covariance parameters will be estimated via the REML method. Compound symmetry will be

used for the covariance structure. The Kenward and Roger's method for evaluating the degrees of freedom will also be used.

The statistical model will be used to calculate the least squares mean (LSMean) treatment difference (elafibranor - placebo), 99% CI, the raw p-value at the end of the 72 week treatment period and the multiplicity adjusted p-value computed according to the pre-defined gatekeeping procedure will be provided. The LSMeans and 99% CI for each treatment group will also be presented.
In addition, the 95% CI will be presented for the LSMean treatment difference and LSMeans for each treatment group.

The analysis of this key secondary endpoint will be performed on the ITT population. Supportive analyses will be performed on PPS population.

7.7.2.2 Other Secondary Efficacy

Unless otherwise stated, the other secondary efficacy endpoints will be summarized for patients in the ITT population.

7.7.2.2.1 Liver biopsy endpoints

Liver biopsy results will be used to summarize histological changes.

Liver biopsy results collected at the end of the 72 week treatment period will be summarized by treatment group for patients in the ITT population.

Change from baseline to end of the 72 week treatment period will be calculated for each liver biopsy endpoint as the value on the liver biopsy report at the end of the 72 week treatment period minus the value of the liver biopsy report at baseline.

7.7.2.2.1.1 Fibrosis (NASH CRN)

The number and percentage of patients with each fibrosis stage based on NASH CRN scoring (Stage 0, Stage 1a, Stage 1b, Stage 1c, Stage 2, Stage 3 and Stage 4) at baseline and at the end of the 72 week treatment period will be summarized by treatment group.

Shift tables presenting the number and percentage of patients with each fibrosis stage according to NASH CRN scoring at baseline compared to the end of the 72 week treatment period according to NASH CRN scoring will be provided for each treatment group. For this summary, Stages 1a, 1b, and 1c will be grouped as Stage 1.

Fibrosis stage according to NASH CRN scoring at baseline, end of the 72 week treatment period, and change from baseline to end of 72 week treatment period will be summarized by treatment group as continuous data, where stages are mapped numerically (e.g. Stage 1a, 1b, and 1c are all set to a value of 1).

The change from baseline in fibrosis stage according to NASH CRN scoring at the end of the 72 week treatment period will be analysed using a non-parametric ANCOVA method of Koch, Tangen, Jung and

Amara (Koch et al.), with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3) and baseline NAS. This analysis will be performed using the SAS macro for non-parametric randomization-based ANCOVA 'NParCov4'. An example of the SAS code for the non-parametric randomization-based ANCOVA is below:

%NParCov4(OUTCOMES=a*val*, COVARS=*diab gender fibrosis NAS*,
HYPOTH=ALT, TRTGRPS=*trtpn*, TRANSFORM=NONE,
DSNIN=<*dataset*>, DSNOUT=<*outdat _outdat_deptest outdat_ci*>);

where

aval = change from baseline analysis value (e.g. NASH CRN at week 72 – baseline score)
trtpn = treatment (numeric value for planned treatment)
diab = type 2 diabetes category (yes, no)
gender = gender category (male, female)
fibrosis = fibrosis stage (F2, F3)
NAS = NAS at baseline (categorical)

Note: The variance under the null (HYPOTH=NULL) will be the structure for producing p-values, while the variance under the alternative (HYPOTH=ALT) will be used for computing confidence intervals.

The statistical model will be used to calculate the treatment difference (elafibranor - placebo), 95% CI and p-value at week 72.

7.7.2.2.1.2 NASH in combination with Fibrosis (NASH CRN)

The number and percentage of patients with improvement of fibrosis of at least one stage according to NASH CRN scoring without worsening of NASH at the end of the 72 week treatment period will be summarized by treatment group. Worsening of NASH is defined as worsening of either steatosis NAS, hepatocyte ballooning NAS or lobular inflammation NAS.

The number of patients with improvement of fibrosis of at least one stage according to NASH CRN scoring without worsening of NASH at the end of the 72 week treatment period will be analysed using the same method as for the primary analysis of the primary endpoint (see section 7.7.1)..

The estimate of difference (elafibranor/placebo) in rate of improvement of fibrosis of at least one stage according to NASH CRN scoring without worsening of NASH, its 95% CI and the corresponding p-value will be presented. The rate of improvement of fibrosis of at least one stage according to NASH CRN scoring without worsening of NASH will also be presented for each treatment group with unadjusted 95% CIs using the Wilson score method.

Patients with missing data for improvement of fibrosis without worsening of NASH will be treated as non-responders for the analysis.

The number and percentage of patients with no worsening of fibrosis of at least one stage according to NASH CRN scoring and no worsening of NASH at the end of the 72 week treatment period will be summarized by treatment group. Worsening of NASH is defined as worsening of either steatosis NAS, hepatocyte ballooning NAS or lobular inflammation NAS.

The number of patients with no worsening of fibrosis of at least one stage according to NASH CRN scoring and no worsening of NASH at the end of the 72 week treatment period will be analysed using the same method as for the primary analysis of the primary endpoint (see section 7.7.1).

The estimate of difference (elafibranor/placebo) in rate of no worsening of fibrosis of at least one stage according to NASH CRN scoring and no worsening of NASH, its 95% CI and the corresponding p-value will be presented. The rate of no worsening of fibrosis of at least one stage according to NASH CRN scoring and no worsening of NASH will also be presented for each treatment group with unadjusted 95% CIs using the Wilson score method.

Patients with missing data for worsening of fibrosis of at least one stage or worsening of NASH will be treated as non-responders for the analysis.

The number and percentage of patients with both a resolution of NASH and improvement of fibrosis of at least one stage according to NASH CRN scoring at the end of the 72 week treatment period will be summarized by treatment group.

The number of patients with both a resolution of NASH and improvement of fibrosis of at least one stage according to NASH CRN scoring at the end of the 72 week treatment period will be analysed using the same method as for the primary analysis of the primary endpoint (see section 7.7.1).

The estimate of difference (elafibranor/placebo) in rate of resolution of NASH and improvement of fibrosis of at least one stage according to NASH CRN scoring, its 95% CI and the corresponding p-value will be presented. The rate of resolution of NASH and improvement of fibrosis of at least one stage according to NASH CRN scoring will also be presented for each treatment group with unadjusted 95% CIs using the Wilson score method.

Patients with missing data for improvement of fibrosis of at least one stage or resolution of NASH will be treated as non-responders for the analysis.

7.7.2.2.1.3 Fibrosis (NAFLD Ishak)

The number and percentage of patients with each fibrosis stage based on NAFLD Ishak scoring (Stage 0, Stage 1, Stage 2, Stage 3, Stage 4, Stage 5 and Stage 6) at baseline and at the end of the 72 week treatment period will be summarized by treatment group.

The number and percentage of patients with an improvement of fibrosis of at least one stage according to NAFLD Ishak scoring at the end of the 72 week treatment period will be summarized by treatment group.

An improvement is defined as a reduction in stage (i.e. a decrease).

The number of patients with an improvement of fibrosis of at least one stage according to NAFLD Ishak scoring at the end of the 72 week treatment period will be analysed using a the same method as for the primary analysis of the primary endpoint (see section 7.7.1).

The estimate of difference (elafibranor/placebo) in rate of improvement of fibrosis of at least one stage according to NAFLD Ishak scoring, its 95% CI and the corresponding p-value will be presented. The rate of improvement of fibrosis of at least one stage according to NAFLD Ishak scoring will also be presented for each treatment group with unadjusted 95% CIs using the Wilson score method.

Patients with missing data for an improvement of at least one stage will be treated as non-responders for the analysis.

Shift tables presenting the number and percentage of patients with each fibrosis stage at baseline compared to the end of the 72 week treatment period according to NASH NAFLD Ishak scoring will be provided for each treatment group.

Fibrosis stage according to NAFLD Ishak scoring at baseline, at the end of the 72 week treatment period, and change from baseline to end of 72 week treatment period will be summarized by treatment group as continuous data.

The change from baseline in NAFLD Ishak scoring at the end of the 72 week treatment period will be analysed using a non-parametric ANCOVA method of Koch et al., with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3) and baseline NAS.

The statistical model will be used to calculate the treatment difference (elafibranor - placebo), 95% CI and p-value at week 72.

7.7.2.2.1.4 NAS

The NAS is the sum of three scores:

- Steatosis NAS (0-3)
- Hepatocyte ballooning NAS (0-2)
- Lobular inflammation NAS (0-3)

The number and percentage of patients with each NAS score at baseline and at the end of the 72 week treatment period will be summarized by treatment group.

The number and percentage of patients with an improvement of NAS score of at least one point and at least two points at the end of the 72 week treatment period will be summarized by treatment group.
In addition, descriptive statistics will be provided by treatment group at the end of the 72 week treatment period for the number and percentage of patients with at least two points improvement of NAS score and at least one point improvement in hepatic ballooning.

An improvement is defined as a reduction in grade (i.e. a decrease).

The number of patients with an improvement in NAS at the end of the 72 week treatment period will be analysed using the same method as for the primary analysis of the primary endpoint (see section 7.7.1).

The estimate of difference (elafibranor/placebo) in rate of improvement of NAS, its 95% CI and the corresponding p-value will be presented. The rate of improvement of NAS will also be presented for each treatment group with unadjusted 95% CIs using the Wilson score method.

The analysis will be performed separately for each improvement category (i.e. at least one point improvement, at least two points improvement, at least two points improvement with at least one point improvement in hepatic ballooning).

Patients with missing data for will be treated as non-responders for the analysis.

Shift tables presenting the number and percentage of patients with each NAS score at baseline compared to the end of the 72 week treatment period will be provided for each treatment group.

NAS score at baseline, at the end of the 72 week treatment period, and change from baseline to end of 72 week treatment period will be summarized by treatment group as continuous data.

The change from baseline in NAS score at the end of the 72 week treatment period will be analysed using non-parametric ANCOVA method of Koch et al., with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3) and baseline NAS score.

The statistical model will be used to calculate the treatment difference (elafibranor - placebo), 95% CI and p-value at week 72.

7.7.2.2.1.5 NAS disease activity score

The NAS disease activity score is the sum of two scores:

- Hepatocyte ballooning NAS (0-2)
- Lobular inflammation NAS (0-3)

The number and percentage of patients with each NAS disease activity score at baseline and at the end of the 72 week treatment period will be summarized by treatment group.

The number and percentage of patients with an improvement of NAS disease activity score of at least one point and at least two points at the end of the 72 week treatment period will be summarized by treatment group.
In addition, descriptive statistics will be provided by treatment group at the end of the 72 week treatment period for the number and percentage of patients with at least one point improvement in NAS disease activity score and with at least one point improvement in hepatic ballooning.

An improvement is defined as a reduction in grade (i.e. a decrease).

The number of patients with an improvement in disease activity score at the end of the 72 week treatment period will be analysed using the same method as for the primary analysis of the primary endpoint (see section 7.7.1). The baseline NAS disease activity score will be introduced into the logistic model as covariate instead of the baseline NAS (categorical).

The estimate of difference (elafibranor/placebo) in rate of improvement of NAS disease activity, its 95% CI and the corresponding p-value will be presented. The rate of improvement of NAS disease activity will also be presented for each treatment group with unadjusted 95% CIs using the Wilson score method.

The analysis will be performed separately for each improvement category (i.e. at least one point improvement, at least one point improvement with at least one point improvement in hepatic ballooning, and at least two points improvement).

Patients with missing data will be treated as non-responders for the analysis.

Shift tables presenting the number and percentage of patients with each NAS disease activity score at baseline compared to the end of the 72 week treatment period will be provided for each treatment group.

NAS disease activity score at baseline, at the end of the 72 week treatment period, and change from baseline to end of 72 week treatment period will be summarized by treatment group as continuous data.

The change from baseline in NAS disease activity score at the end of the 72 week treatment period will be analysed using non-parametric ANCOVA method of Koch et al., with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3) and baseline NAS disease activity score.

The statistical model will be used to calculate the treatment difference (elafibranor - placebo), 95% CI and p-value at week 72.

7.7.2.2.1.6 SAF score

The Steatosis-Activity-Fibrosis (SAF) score is comprised of three separate scores:

- Steatosis (S) with grades from 0 to 3;
- Activity* (A) with grades from 0 to 4;
- Fibrosis (F) with grades from 0 to 4;

*SAF activity score is the un-weighted addition of hepatocyte ballooning (0-2) and lobular inflammation (0-2).

The number and percentage of patients with each SAF activity score at baseline, and at the end of the 72 week treatment period will be summarized by treatment group.

The number and percentage of patients with an improvement in SAF activity score of at least one point or at least two points at the end of the 72 week treatment period will be summarized by treatment group.

In addition, descriptive statistics will be provided by treatment group at the end of the 72 week treatment period for the number and percentage of patients with at least one point improvement in SAF activity score and with at least one point improvement in hepatic ballooning,

An improvement is defined as a reduction in score (i.e. a decrease).

The number of patients with an improvement in SAF activity score at the end of the 72 week treatment period will be analysed using the same method as for the primary analysis of the primary endpoint (see section 7.7.1). The baseline SAF activity score will be introduced into the logistic model as covariate instead of the baseline NAS (categorical).

The estimate of difference (elafibranor/placebo) in rate of improvement of SAF activity score, its 95% CI and the corresponding p-value will be presented. The rate of improvement of SAF activity score will also be presented for each treatment group with unadjusted 95% CIs using the Wilson score method.

The analysis will be performed separately for each improvement category (i.e. at least one point improvement, at least one point improvement with at least one point improvement in hepatic ballooning, and at least two points improvement).

Patients with missing data will be treated as non-responders for the analysis.

Shift tables presenting the number and percentage of patients with each SAF activity score at baseline compared to the end of the 72 week treatment period will be provided for each treatment group.

SAF activity score at baseline, at the end of the 72 week treatment period and change from baseline to end of 72 week treatment period will be summarized by treatment group as continuous data.

The change from baseline in SAF activity score at the end of the 72 week treatment period will be analysed using non-parametric ANCOVA method of Koch et al., with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3) and baseline SAF activity score.

The statistical model will be used to calculate the treatment difference (elafibranor - placebo), 95% CI and p-value at week 72.

7.7.2.2.1.7 Steatosis (NAS/SAF)

The number and percentage of patients with each steatosis grade based on NAS/SAF scoring (Grade 0, Grade 1, Grade 2 and Grade 3) at baseline and at the end of the 72 week treatment period will be summarized by treatment group.

The number and percentage of patients with an improvement of steatosis of at least one grade at the end of the end of the 72 week treatment period will be summarized by treatment group.

An improvement is defined as a reduction in grade (i.e. a decrease).

The number of patients with an improvement of steatosis of at least one grade at the end of the 72 week treatment period will be analysed using the same method as for the primary analysis of the primary endpoint (see section 7.7.1). The baseline steatosis grade based on NAS/SAF scoring will be introduced into the logistic model as covariate instead of the baseline NAS (categorical).

The estimate of difference (elafibranor/placebo) in rate of improvement of steatosis of at least one grade, its 95% CI and the corresponding p-value will be presented. The rate of improvement of steatosis of at least one grade will also be presented for each treatment group with unadjusted 95% CIs using the Wilson score method.

Patients with missing data for an improvement of at least one grade will be treated as non-responders for the analysis.

Shift tables presenting the number and percentage of patients with each steatosis grade according to NAS/SAF scoring at baseline compared to the end of the 72 week treatment period will be provided for each treatment group.

Steatosis grading according to NAS/SAF scoring scores at baseline, at the end of the 72 week treatment period, and change from baseline to end of 72 week treatment period will be summarized by treatment group as continuous data.

The change from baseline in steatosis grading according to NAS/SAF scoring at the end of the 72 week treatment period will be analysed using a non-parametric ANCOVA method of Koch et al., with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3) and baseline steatosis grading according to NAS/SAF scoring.

The statistical model will be used to calculate the treatment difference (elafibranor - placebo), 95% CI and p-value at week 72.

7.7.2.2.1.8 Hepatic Ballooning (NAS/SAF)

The number and percentage of patients with each hepatic ballooning grade based on NAS/SAF scoring (Grade 0, Grade 1 and Grade 2) at baseline and at the end of the 72 week treatment period will be summarized by treatment group.

The number and percentage of patients with an improvement of hepatic ballooning of at least one grade at the end of the 72 week treatment period will be summarized by treatment group.

An improvement is defined as a reduction in grade (i.e. a decrease).

The number of patients with an improvement of hepatic ballooning of at least one grade at the end of the 72 week treatment period will be analysed using the same method as for the primary analysis of the primary endpoint (see section 7.7.1). The baseline hepatic ballooning grade based on NAS/SAF scoring will be introduced into the logistic model as covariate instead of the baseline NAS (categorical).

The estimate of difference (elafibranor/placebo) in rate of improvement of hepatic ballooning of at least one grade, its 95% CI and the corresponding p-value will be presented. The rate of improvement of hepatic ballooning of at least one grade will also be presented for each treatment group with unadjusted 95% CIs using the Wilson score method.

Patients with missing data for an improvement of at least one grade will be treated as non-responders for the analysis.

Shift tables presenting the number and percentage of patients with each hepatic ballooning grade at baseline compared to the end of the 72 week treatment period will be provided for each treatment group.

Hepatic ballooning grades at baseline, at the end of the 72 week treatment period, and change from baseline to end of 72 week treatment period will be summarized by treatment group as continuous data.

The change from baseline in hepatic ballooning grading according to NAS/SAF scoring at the end of the 72 week treatment period will be analysed using a non-parametric ANCOVA method of Koch et al., with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3) and baseline hepatic ballooning grade based on NAS/SAF scoring.

The statistical model will be used to calculate the treatment difference (elafibranor - placebo), 95% CI and p-value at week 72.

7.7.2.2.1.9 Lobular inflammation (NAS)

The number and percentage of patients with each lobular inflammation grade based on NAS scoring (Grade 0, Grade 1, Grade 2 and Grade 3) at baseline and at the end of the 72 week treatment period will be summarized by treatment group.

The number and percentage of patients with an improvement of lobular inflammation of at least one grade based on NAS scoring at the end of the 72 week treatment period will be summarized by treatment group.

An improvement is defined as a reduction in grade (i.e. a decrease).

The number of patients with an improvement of lobular inflammation of at least one grade based on NAS scoring at the end of the 72 week treatment period will be analysed using the same method as for the primary analysis of the primary endpoint (see section 7.7.1). The baseline lobular inflammation grade based on NAS scoring will be introduced into the logistic model as covariate instead of the baseline NAS (categorical).

The estimate of difference (elafibranor/placebo) in rate of improvement of lobular inflammation of at least one grade, its 95% CI and the corresponding p-value will be presented. The rate of improvement of lobular inflammation of at least one grade based on NAS scoring will also be presented for each treatment group with unadjusted 95% CIs using the Wilson score method.

Patients with missing data for an improvement of at least one grade will be treated as non-responders for the analysis.

Shift tables presenting the number and percentage of patients with each lobular inflammation grade based on NAS scoring at baseline compared to the end of the 72 week treatment period will be provided for each treatment group.

Lobular inflammation grades based on NAS scoring at baseline, at the end of the 72 week treatment period, and change from baseline to end of 72 week treatment period will be summarized by treatment group as continuous data.

The change from baseline in lobular inflammation grading based on NAS scoring at the end of the 72 week treatment period will be analysed using a non-parametric ANCOVA method of Koch et al., with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3) and baseline lobular inflammation grade based on NAS scoring.

The statistical model will be used to calculate the treatment difference (elafibranor - placebo), 95% CI and p-value at week 72.

7.7.2.2.1.10 Lobular inflammation (SAF)

The number and percentage of patients with each lobular inflammation grade based on SAF scoring (Grade 0, Grade 1 and Grade 2) at baseline and at the end of the 72 week treatment period will be summarized by treatment group.

The number and percentage of patients with an improvement of lobular inflammation of at least one grade based on SAF scoring at the end of the 72 week treatment period will be summarized by treatment group.

An improvement is defined as a reduction in grade (i.e. a decrease).

The number of patients with an improvement of lobular inflammation of at least one grade based on SAF scoring at the end of the 72 week treatment period will be analysed using the same method as for the primary analysis of the primary endpoint (see section 7.7.1). The baseline lobular inflammation grade based on SAF scoring will be introduced into the logistic model as covariate instead of the baseline NAS (categorical).

The estimate of difference (elafibranor/placebo) in rate of improvement of lobular inflammation of at least one grade based on SAF scoring, its 95% CI and the corresponding p-value will be presented. The rate of improvement of lobular inflammation of at least one grade based on SAF scoring will also be presented for each treatment group with unadjusted 95% CIs using the Wilson score method.

Patients with missing data for an improvement of at least one grade will be treated as non-responders for the analysis.

Shift tables presenting the number and percentage of patients with each lobular inflammation grade based on SAF scoring at baseline compared to the end of the 72 week treatment period will be provided for each treatment group.

Lobular inflammation grades based on SAF scoring at baseline, at the end of the 72 week treatment period, and change from baseline to end of 72 week treatment period will be summarized by treatment group as continuous data.

The change from baseline in lobular inflammation grading based on SAF scoring at the end of the 72 week treatment period will be analysed using non-parametric ANCOVA method of Koch et al., with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3) and baseline lobular inflammation grade based on SAF scoring.

The statistical model will be used to calculate the treatment difference (elafibranor - placebo), 95% CI and p-value at week 72.

7.7.2.2.1.11 Portal inflammation

The number and percentage of patients with each portal inflammation grade (Grade 0, Grade 1 and Grade 2) at baseline and at the end of the 72 week treatment period will be summarized by treatment group.

The number and percentage of patients with an improvement of portal inflammation of at least one grade at the end of the 72 week treatment period will be summarized by treatment group.

An improvement is defined as a reduction in grade (i.e. a decrease).

The number of patients with an improvement of portal inflammation of at least one grade at the end of the 72 week treatment period will be analysed using the same method as for the primary analysis of the primary endpoint (see section 7.7.1). The baseline portal inflammation grade will be introduced into the logistic model as covariate instead of the baseline NAS (categorical).

The estimate of difference (elafibranor/placebo) in rate of improvement of portal inflammation, its 95% CI and the corresponding p-value will be presented. The rate of improvement of portal inflammation will also be presented for each treatment group with unadjusted 95% CIs using the Wilson score method.

Patients with missing data for an improvement of at least one grade will be treated as non-responders for the analysis.

Shift tables presenting the number and percentage of patients with each portal inflammation grade at baseline compared to the end of the 72 week treatment period will be provided for each treatment group.

Portal inflammation grades at baseline, at the end of the 72 week treatment period, and change from baseline to end of 72 week treatment period will be summarized by treatment group as continuous data.

The change from baseline in portal inflammation at the end of the 72 week treatment period will be analysed using non-parametric ANCOVA method of Koch et al., with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3) and baseline portal inflammation grade.

The statistical model will be used to calculate the treatment difference (elafibranor - placebo), 95% CI and p-value at week 72.

7.7.2.2.2 Area of fibrosis assessed by morphometry

Area of fibrosis (normalized Collagen Proportional are in %) assessed by morphometry at baseline, at the end of the 72 week treatment period, and change from baseline to end of 72 week treatment period will be summarized by treatment group as continuous data.

The change from baseline in area of fibrosis at the end of the 72 week treatment period will be analysed using non-parametric ANCOVA method of Koch et al., with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3) and baseline area of fibrosis.

The statistical model will be used to calculate the treatment difference (elafibranor - placebo), 95% CI and p-value at week 72.

7.7.2.2.3 Liver enzymes and liver markers

Liver enzymes and liver markers to be summarized include Albumin, ALT, aspartate aminotransferase (AST), Total and Direct Bilirubin, GGT, Protein total, and Alkaline Phosphatase.

Liver enzyme and liver marker values recorded at each time-point (baseline, week 12, week 24, week 36, week 48, week 60, week 72, and at the end of the 72 week treatment period) and change from baseline to each post baseline time-point will be summarized by treatment group and endpoint as continuous data.

Line plots of means with whiskers extending to the mean +/ SE over time using the same visits as for table summaries will be presented for ALT, AST, AST/ALT ratio, GGT and Alkaline Phosphatase.

The change from baseline in liver enzyme and liver marker values at each time-point will be analysed using a repeated measures ANCOVA model with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3), endpoint specific liver enzyme/liver marker baseline value, time-point and treatment by time-point interaction. Covariance parameters will be estimated via the REML method. Compound symmetry will be used for the covariance structure. The Kenward and Roger's method for evaluating the degrees of freedom will also be used.

The analysis will be performed using 'PROC MIXED'. An example of the SAS code is below:

```
proc mixed data = <dataset> method=reml;
  class trtpn avisit diab gender fibrosis usubjid;
  model aval = trtpn diab gender fibrosis base trtpn*avisit / ddfm = KR;
  repeated avisit / subject = usubjid type = CS;
  lsmeans trtpn*avisit / diff cl;
run;
```

where

aval = value of endpoint to be analysed (e.g. change from baseline in endpoint specific liver enzyme/liver marker value)

avisit = time-point (visit)
trtpn = treatment (numeric value for planned treatment)
diab = type 2 diabetes category (yes, no)
gender = gender category (male, female)
fibrosis = fibrosis stage (F2, F3)
base = endpoint specific liver enzyme/liver marker baseline value (continuous)
usubjid = Unique subject identifier

The statistical model will be used to calculate the least squares mean (LSMean) treatment difference (elafibranor - placebo) and 95 CI% at the end of the 72 week treatment period. The LSMeans and 95% CI for each treatment group, LSMean treatment difference, 95% CI and corresponding p-value will also be presented at each time-point.

7.7.2.2.4 Non-invasive markers of fibrosis and steatosis

Non-invasive markers of fibrosis and steatosis to be summarized include CK18 (M65 & M30), adiponectin, ferritin, FGF19 & FGF21, alpha2 macroglobulin, hyaluronic acid, PIIINP, TIMP-1, CHI3L1, Fibrotest, ELF, NAFLD Fibrosis score, Steatotest, FLI, Fibrometre S, and FIB-4.

Non-invasive markers of fibrosis and steatosis values recorded at each time-point (baseline, week 12, week 24, week 36, week 48, week 60, week 72, and at the end of the 72 week treatment period) and change from baseline to each post baseline time-point will be summarized by treatment group and endpoint as continuous data.

Line plots of means with whiskers extending to the mean +/ SE over time using the same visits as for table summaries will be presented for CK18 (M65 & M30), adiponectin, ferritin, FGF19 & FGF21, alpha2 macroglobulin, hyaluronic acid, PIIINP, and TIMP-1.

The change from baseline in non-invasive markers of fibrosis and steatosis values at each time-point will be analysed using a repeated measures ANCOVA model with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3), baseline endpoint specific non-invasive marker of fibrosis and steatosis value, time-point, and treatment by time-point interaction. Covariance parameters will be estimated via the REML method. Compound symmetry will be used for the covariance structure. The Kenward and Roger's method for evaluating the degrees of freedom will also be used.

The statistical model will be used to calculate the least squares mean (LSMean) treatment difference (elafibranor - placebo) and 95 CI% at the end of the 72 week treatment period. The LSMeans and 95% CI for each treatment group, LSMean treatment difference, 95% CI and corresponding p-value will also be presented at each time-point.

7.7.2.2.5 Lipid parameters

Lipid parameters to be summarized include total triglycerides, total cholesterol and fractions, small dense LDL, ApoAII, Apo CIII, and Apo E and non-esterified fatty acids.

Lipid parameter values recorded at each time-point (baseline, week 24, week 36, week 48, week 60, week 72, and at the end of the 72 week treatment period) and change from baseline to each post baseline time-point will be summarized by treatment group and endpoint as continuous data.

In addition, relative change (percentage change from baseline) will also be summarized as described above.

Line plots of means with whiskers extending to the mean +/ SE over time using the same visits as for table summaries will be presented for total triglycerides, total cholesterol and fractions, and Apo E.

The change from baseline in lipid parameter values at each time-point will be analysed using a repeated measures ANCOVA model with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3), baseline endpoint specific lipid parameter value, time-point, and treatment by time-point interaction. Covariance parameters will be estimated via the REML method. Compound symmetry will be used for the covariance structure. The Kenward and Roger's method for evaluating the degrees of freedom will also be used.

The statistical model will be used to calculate the least squares mean (LSMean) treatment difference (elafibranor - placebo) and 95 CI% at the end of the 72 week treatment period. The LSMeans and 95% CI for each treatment group, LSMean treatment difference, 95% CI and corresponding p-value will also be presented at each time-point.

In addition, the relative change from baseline in lipid parameters at each time-point will be analysed as described above.

7.7.2.2.6 Body weight

Body weight (kg) recorded at each time-point (baseline, week 12, week 24, week 36, week 48, week 60, week 72, and at the end of the 72 week treatment period) and change from baseline to each post baseline time-point will be summarized by treatment group and endpoint as continuous data.

The change from baseline in body weight at each time-point will be analysed using a repeated measures ANCOVA model with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3), baseline body weight, time-point, and treatment by time-point interaction. Covariance parameters will be estimated via the REML method. Compound symmetry will be used for the covariance structure. The Kenward and Roger's method for evaluating the degrees of freedom will be used.

The statistical model will be used to calculate the LSMean treatment difference (elafibranor - placebo) and 95 CI% at the end of the 72 week treatment period. The LSMeans and 95% CI for each treatment group, LSMean treatment difference, 95% CI and corresponding p-value will also be presented at each time-point.

7.7.2.2.7 Insulin resistance and glucose homeostasis markers

Insulin resistance and glucose homeostasis markers to be summarized include HbA1c, glucose, insulin, c-peptide, fructosamine and HOMA-IR. The below summaries and analysis of HbA1c will be performed on the subgroup of diabetic patients.

Insulin resistance and glucose homeostasis marker values recorded at each time-point (baseline, week 24, week 48, week 72, and at the end of the 72 week treatment period) and change from baseline to each post baseline time-point will be summarized by treatment group and endpoint as continuous data.

In addition, relative change (percentage change from baseline) will also be summarized as described above.

A line plot of means with whiskers extending to the mean +/ SE over time using the same visits as for table summaries will be presented for HOMA-IR.

The change from baseline in insulin resistance and glucose homeostasis marker values at each time-point will be analysed using a repeated measures ANCOVA model with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3), baseline endpoint specific insulin resistance and glucose homeostasis marker value, time-point, and treatment by time-point interaction. Covariance parameters will be estimated via the REML method. Compound symmetry will be used for the covariance structure. The Kenward and Roger's method for evaluating the degrees of freedom will also be used.

The statistical model will be used to calculate the LSMean treatment difference (elafibranor - placebo) and 95 CI% at the end of the 72 week treatment period. The LSMeans and 95% CI for each treatment group, LSMean treatment difference, 95% CI and corresponding p-value will also be presented at each time-point.

7.7.2.2.8 Inflammatory markers

Inflammatory markers to be summarized include hsCRP, haptoglobin and fibrinogen. hsCRP will be log-transformed for the purpose of summaries and analysis.

Inflammatory marker values recorded at each time-point (baseline, week 24, week 48, week 72, and at the end of the 72 week treatment period) and change from baseline to each post baseline time-point will be summarized by treatment group and endpoint as continuous data.

Line plots of means with whiskers extending to the mean +/ SE over time using the same visits as for table summaries will be presented for haptoglobin and fibrinogen.

The change from baseline in inflammatory markers at each time-point will be analysed using a repeated measures ANCOVA model with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3), baseline endpoint specific inflammatory marker baseline value, time-point, and treatment by time-point interaction. Covariance parameters will be estimated via the REML method. Compound symmetry will be used for the covariance structure. Compound symmetry will be used for the covariance structure. The Kenward and Roger's method for evaluating the degrees of freedom will also be used.

The statistical model will be used to calculate the LSMean treatment difference (elafibranor - placebo) and 95 CI% at the end of the 72 week treatment period. The LSMeans and 95% CI for each treatment group, LSMean treatment difference, 95% CI and corresponding p-value will also be presented at each time-point.

7.7.2.2.9 Cardiovascular risk profile

Framingham Risk Score provides an estimate of a patient's 10 year risk of developing cardiovascular disease.

Framingham Risk Score will be calculated based on the following information recorded at each visit: age, gender, total cholesterol (mmol/L), HDL cholesterol (mmol/L), smoker (yes, no), diabetes (yes, no), systolic blood pressure (mmHg) and is the patient being treated for high blood pressure (yes, no).

For each patient, the Framingham Risk Scores will be calculated based on the publication by D'Agostino et al (2008) as follows:

$$\text{Risk Score} = 100 * (1 - \text{RiskPeriodFactor}^{e(\text{RiskFactors})})$$

RiskFactors = (ln(Age) * AgeFactor) + (ln(TotalChol) * TotalCholFactor) + (ln(HDLChol) * HDLCholFactor) + (ln(SysBP) * SysBPFactor) + Smoking + Diabetes - AvgRisk

where

For women: **Age Factor** = 2.32888; **Total Chol Factor** = 1.20904; **HDL Chol Factor** = -0.70833; **SBP Risk Factor if not treated** = 2.76157; **SBP Risk Factor if treated** = 2.82263; **Smoking**=0.52873; **Diabetes**=0.69154; **Avg Risk** = 26.1931 and **Risk Period Factor** = 0.95012

For men: **Age Factor** = 3.06117; **Total Chol Factor** = 1.12370; **HDL Chol Factor** = -0.93263; **SBP Risk Factor if not treated** = 1.93303; **SBP Risk Factor if treated** = 1.99881; **Smoking**=0.65451; **Diabetes**=0.57367; **Avg Risk** = 23.9802 and **Risk Period Factor** = 0.88936

Framingham Risk Scores calculated at baseline, at the end of the 72 week treatment period, and change from baseline to end of 72 week treatment period will be summarized by treatment group as continuous data.

The change from baseline in Framingham Risk Scores at the end of the 72 week treatment period will be analysed using a non-parametric ANCOVA method of Koch et al., with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3) and baseline Framingham Risk score.

The statistical model will be used to calculate the treatment difference (elafibranor - placebo), 95% CI and p-value at week 72.

7.7.2.2.10 Quality of life

Quality of life will be assessed using the SF-36 questionnaire.

Scores for each of the eight health scale domains (Physical Functioning, Role-Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role-Emotional and Mental Health) and the two component summary measures (Physical Component Summary and Mental Component Summary) will be calculated for each patient at each time-point as described in the SF-36 scoring manual (please see Appendix III for further details).

Scores for each of the eight health scale domains and the two component summary measures at each visit (baseline, week 24, week 48, week 72, and at the end of the 72 week treatment period) and change from baseline to each post-baseline visit will be summarized by treatment group as continuous data.

The change from baseline in each of the eight health scale domains and the two component summary measures at each time-point will be analysed using a repeated measures ANCOVA model with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3), baseline endpoint specific SF-36 score, time-point, and treatment by time-point interaction. Covariance parameters will be estimated via the REML method. Compound symmetry will be used for the covariance structure. The Kenward and Roger's method for evaluating the degrees of freedom will also be used.

Note: The statistical analysis of each of the eight health scale domains and the two component summary measures will be performed separately.

The statistical model will be used to calculate the LSMean treatment difference (elafibranor - placebo) and 95 CI% at the end of the 72 week treatment period. The LSMeans and 95% CI for each treatment group, LSMean treatment difference, 95% CI and corresponding p-value will also be presented at each time-point.

7.7.2.2.11 Liver stiffness by FibroScan measurement

Liver stiffness is measured throughout the study. Liver stiffness measures (LSM) (median value) at each time-point (baseline and week 72) will be summarized by treatment group and endpoint as continuous data, alongside change from baseline for post-baseline assessments.

The change from baseline in LSM at the end of the 72 week treatment period will be analysed using a non-parametric ANCOVA method of Koch et al., with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3) and baseline LSM.

The statistical model will be used to calculate the treatment difference (elafibranor - placebo), 95% CI and p-value at week 72.

7.7.3 Supplementary and Sensitivity Analysis

7.7.3.1 Resolution of NASH without worsening of fibrosis

Statistical Model

A sensitivity analysis to the primary analysis model will be performed. The number of patients with resolution of NASH without worsening of fibrosis will be compared between groups using the Cochran-

Mantel-Haenszel (CMH) test stratified by the randomization strata (type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3) and baseline NAS (=4, >4).
The estimate of the difference (elafibranor/placebo) in rate of resolution of NASH without worsening of fibrosis and its 99% confidence interval and the corresponding p-value will be provided.

This sensitivity analysis will performed using the approach and SAS code described in Stokes et al. (2012), page 50-53.

The p-value of the CMH test will be provided using 'PROC FREQ'. An example of the SAS code is below:

proc freq data = <*dataset*>;
tables *diab*gender*fibrosis*NAS * trtpn * aval* / cmh alpha=0.01 ;
run;

where

aval = flag for event, where event (1), and no event (0)
trtpn = treatment (numeric value for planned treatment)
diab = type 2 diabetes category (yes, no)
gender = gender category (male, female)
fibrosis = fibrosis stage (F2, F3)
NAS = NAS at baseline (=4, >4)

In addition, the difference (elafibranor/placebo) in rate of resolution of NASH without worsening of fibrosis and its 99% confidence interval will be provided.

This difference will be measured using a weighted average of the difference of proportions across the strata, and will be produced using 'PROC GLM'. An example of the SAS code is below:

proc glm data = <*dataset*>;
class diab gender fibrosis NAS trtpn;
model aval = diab gender fibrosis NAS trtpn;
estimate 'direction' treatment -1 1;
run;

where

aval = flag for event, where event (1), and no event (0)
trtpn = treatment (numeric value for planned treatment)
diab = type 2 diabetes category (yes, no)
gender = gender category (male, female)
fibrosis = fibrosis stage (F2, F3)
NAS = NAS at baseline (=4, >4)

The 99% confidence interval will be constructed based on the binomial distribution by including a continuity correction for the variance.

The analysis will be performed on the ITT population.

Missing Data

Patients with missing data for resolution of NASH without worsening of fibrosis will be treated as a non-responder for the main analysis.

Supplementary and sensitivity analysis of resolution of NASH without worsening of fibrosis will be performed using multiple imputations and a pattern mixture model will be performed.

Firstly, missing values for resolution of NASH without worsening of fibrosis will be imputed multiple times to account for the uncertainty about the true values to impute. Missing values are assumed to be monotone in nature, i.e. that the covariates will be present for all subjects, and the response value will be imputed using the monotone logistic regression method. If the monotone assumption does not hold, other imputation methods will be deployed. For this multiple imputation analysis, a separate random seed between 1 and 10000 will be generated. The random seed to be used for the imputation will be 8993. One hundred imputed datasets will be generated.

An example of the SAS code to be used for the multiple imputation method using the monotone approach is below:

proc mi data = <*dataset*> seed=<seed>
nimpute = *100;*
class trtpn response;
var trtpn diab gender fibrosis NAS response;
monotone logistic (response =trtpn diab gender fibrosis NAS);
run;

where

response = responder(1), non-responder (0)
trtpn = treatment (numeric value for planned treatment)
diab = type 2 diabetes category (yes, no)
gender = gender category (male, female)
fibrosis = fibrosis stage (F2, F3)
NAS = NAS at baseline (categorical)

The one hundred imputed datasets will be used separately in the statistical analysis model (see section 7.7.1) and the resulting estimate of difference (elafibranor/placebo) in rate of resolution of NASH without worsening of fibrosis and its 99% CIs will be combined via Rubin's rules (using PROC MIANALYZE) to provide a global estimate of the difference with a corresponding 99% CI and p-value.

Secondly, missing values for the surrogate endpoint (resolution of NASH without worsening of fibrosis) will be imputed using a pattern mixture model. This analysis will use the placebo group as the basis for imputation of any missing values after the last available data point for a patient. To implement this approach, the MNAR statement of PROC MI will be used along with the monotone logistic regression (similarly to the imputation strategy described above).

An example of the SAS code to be used for the placebo-based multiple imputation method using the monotone approach is provided below:

```
proc mi data = <dataset> seed=<seed> out=<output> nimpute = 100;
class trtpn response;
var diab gender fibrosis NAS response;
monotone logistic (response = diab gender fibrosis NAS);
mnar model (response / modelobs=( trtpn ='1')); /* Here trtpn ='1' means placebo.*/
run;
```

The proposed pattern-mixture analysis imputes missing values for the surrogate endpoint using only the selected baseline variables.

The analysis will be performed on the ITT population.

Center

Sensitivity analysis will be performed to explore the impact of center.

The number of patients with resolution of NASH without worsening of fibrosis will be analysed using the same method as in the primary analysis of the primary endpoint (see section 7.7.1) with the addition of center as a covariate into the logistic model.

Firstly, the center will be introduced into the model as fixed effect. In case of centers with less than 10 patients, they will be pooled with other centers in their country. If pooling by country results in pooled centers with fewer than 10 patients, the pooled center will be combined with the center within the country with the fewest number of patients of 10 or more. In the event of a tie in the number of patients for pooling, the lower value of Site ID will be preferred.

If after performing pooling by country within a given country does not result in a pooled center of greater than or equal to 10 patients, the following countries will have their pooled centers combined:

- Argentina and Chile
- Belgium and Netherlands
- Denmark, Finland and Sweden
- France and Switzerland
- Russia and Turkey
- Spain and Portugal

Secondly, the center (without pooling) will be introduced into the model as random effect. This will be accomplished using generalized linear models with random effects implemented in PROC GLIMMIX, with a random intercept and a center class variable included as subject in the RANDOM statement (RANDOM INT/SUBJECT=INV). The binomial error distribution and logit link need to be specified in the model statement. Therefore the fitted model is

$$logit\big(P(y_i = 1)\big) = \beta_0 + \beta_1 t_i + x_i^T \beta_2 + b_{l(i)},$$

where $l(i)$ denotes the center number where the i^{th} patient belongs, and $b_{l(i)} \sim N(0, \sigma_b^2)$ is the random effect with variance σ_b^2.

The solution for random effects can be obtained with estimated response rates defined as

$$p_{i,0} = P_{\hat{\beta},\hat{b}_{l(i)}}(Y = 1 | x = x_i, t = 0) = \left[1 + \exp\left(-\hat{\beta}_0 - x_i^T \hat{\beta}_2 - \hat{b}_{l(i)}\right)\right]^{-1}$$

and

$$p_{i,1} = P_{\hat{\beta},\hat{b}_{l(i)}}(Y = 1 | x = x_i, t = 1) = \left[1 + \exp\left(-\hat{\beta}_0 - \hat{\beta}_1 - x_i^T \hat{\beta}_2 - \hat{b}_{l(i)}\right)\right]^{-1},$$

where $\hat{b}_{l(i)}$ is the predicted random effect associated with the center $l(i)$ where the ith patient belongs. Define vectors A_1 and A_0 with elements $a_{i0} = p_{i0}(1 - p_{i0})$ and $a_{i1} = p_{i1}(1 - p_{i1})$ as before. Create enhanced matrices $\check{X}_0$ and $\check{X}_1$ by appending to $\tilde{X}_0$ and $\tilde{X}_1$ the last column of 1's.

To account for variability due to random effects for patients from the same center, define the center-specific variance matrices $V_{l(i)}$, that enhance the coefficient covariance matrix V by appending an extra row and an extra column containing in 0's for off-diagonal elements and $\hat{\sigma}_b^2 + \widehat{var}\left(\hat{b}_{l(i)}\right)$ at the lower diagonal element: a sum of the estimated variance of the random effect $\hat{\sigma}_b^2$ and the estimated variance of the best unbiased prediction of a specific center's random effect $\widehat{var}\left(\hat{b}_{l(i)}\right)$. The variance of the estimator can be constructed as

$$var(d) = \frac{1}{N^2}\left\{\sum_{l=1}^{L} \sum_{i,j:l(i)=l(j)=l} \left(a_{i0}a_{j0}\check{x}_{i0}^T V_{l(i)} \check{x}_{j0} + a_{i1}a_{j1}\check{x}_{i1}^T V_{l(i)} \check{x}_{j1} + a_{i1}a_{j0}\check{x}_{i1}^T V_{l(i)} \check{x}_{j0}\right) + \sum_{i,j:l(i)\neq l(j)} \left(a_{i0}a_{j0}\tilde{x}_{i0}^T V \tilde{x}_{j0} + a_{i1}a_{j1}\tilde{x}_{i1}^T V \tilde{x}_{j1} + a_{i1}a_{j0}\tilde{x}_{i1}^T V \tilde{x}_{j0}\right)\right\},$$

where L is the number of centers.

The p-value will be computed using the normal approximation method.

The analyses will be performed on the ITT population.

7.7.4 Subgroup Analysis

Exploratory analyses of the primary and key secondary endpoints will be performed for the following subgroups:

- Presence of type 2 diabetes (yes, no)
- Gender (male, female)

- Baseline fibrosis (F1, F2, F3)
- Geographical region (North America, Europe, South America, Rest of World)
- Race (Caucasian, Other)
- Ethnicity (Hispanic, Not Hispanic)
- Age (< 60 years, ≥ 60 years)
- PNPLA3 (absence or presence of risk allele G (i.e. CC vs GG/CG), and within the at risk group, homozygous vs. heterozygous for the risk allel G (i.e. CG vs GG)
- Patients under statins (yes/no), defined as patients who have duration of IP exposure greater or equal to 60 days, with extent of exposure to statin greater than or equal to 30 days before the visit 7 biopsy date.

Unless otherwise stated, the subgroup analysis will be performed on patients in the ITT population. Analyses using fibrosis subgroups will be performed on the FITT population.

7.7.4.1 Resolution of NASH without worsening of fibrosis

The number and percentage of patients with resolution of NASH without worsening of fibrosis will be summarized by treatment group for each subgroup for patients in the ITT population.

Patients with missing data for resolution of NASH without worsening of fibrosis will be treated as non-responders for the main analysis.

For each subgroup, the number of patients with resolution of NASH without worsening of fibrosis will be analysed using the same method as for the primary analysis of the primary endpoint (see section 7.7.1) with a logistic regression model with a fixed term for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3), baseline NAS and treatment by subgroup interaction used for generating the probability of response per patient.

For each subgroup, the estimate of difference (elafibranor/placebo) in rate of resolution of NASH without worsening of fibrosis and its 95% CI will be presented. The rate of resolution of NASH without worsening of fibrosis will also be presented for each treatment group and subgroup with unadjusted 95% CIs using the Wilson score method.

Note: For the subgroup analysis by presence of type 2 diabetes the fixed term for type 2 diabetes will be removed. For the subgroup analysis by gender the fixed term for gender will be removed. For the subgroup analysis by fibrosis stage the fixed term for fibrosis stage will be removed.

The estimate of difference (elafibranor/placebo) and 95% CI will be presented for each subgroup on a forest plot.

7.7.4.2 Improvement in fibrosis according to NASH CRN scoring

The number and percentage of patients with improvement of fibrosis according to NASH CRN scoring will be summarized by treatment group for each subgroup.

Patients with missing data improvement in fibrosis according to NASH CRN scoring will be treated as non-responders for the main analysis.

For each subgroup, the number of patients with improvement in fibrosis according to NASH CRN scoring will be analysed using the same method as for the primary analysis of the primary endpoint (see section 7.7.1) with a logistic regression model with a fixed term for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3), baseline NAS and treatment by subgroup interaction used for generating the probability of response per patient.

For each subgroup, the estimate of difference (elafibranor/placebo) in rate of improvement in fibrosis according to NASH CRN scoring and its 95% CI will be presented. The rate of improvement in fibrosis according to NASH CRN scoring will also be presented for each treatment group and subgroup with unadjusted 95% CIs using the Wilson score method.

Note: For the subgroup analysis by presence of type 2 diabetes the fixed term for type 2 diabetes will be removed. For the subgroup analysis by gender the fixed term for gender will be removed. For the subgroup analysis by fibrosis stage the fixed term for fibrosis stage will be removed.

The estimate of difference (elafibranor/placebo) and 95% CI for each subgroup will be presented on a forest plot.

7.7.4.3 Metabolic parameters

For each subgroup, the metabolic parameter values recorded at each time-point (baseline, week 24, week 36, week 48, week 60, week 72, and at the end of the 72 week treatment period) and change from baseline to each post baseline time-point will be summarized by treatment group and endpoint as continuous data.

In addition, relative change (percentage change from baseline) will also be summarized as described above.

The change from baseline in metabolic parameter values at each time-point will be analysed using a repeated measures ANCOVA model with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F2, F3), baseline endpoint specific parameter value, time-point, and subgroup-by-treatment-by-visit interaction term (as well as subgroup-by-treatment and subgroup-by-visit interaction terms). Covariance parameters will be estimated via the REML method. Compound symmetry will be used for the covariance structure. The Kenward and Roger's method for evaluating the degrees of freedom will also be used.

The statistical model will be used to calculate the least squares mean (LSMean) treatment difference (elafibranor - placebo) and 95 CI% at the end of the 72 week treatment period. The LSMeans and 95% CI for each treatment group, LSMean treatment difference, 95% CI and corresponding p-value will also be presented at each time-point.

Note: For the subgroup analysis by presence of type 2 diabetes the fixed term for type 2 diabetes will be removed. For the subgroup analysis by gender the fixed term for gender will be removed. For the subgroup analysis by fibrosis stage the fixed term for fibrosis stage will be removed.

The LSMean treatment difference and 95% CI for each subgroup (and overall) will be presented at Week 72 on a forest plot. The p-value for the subgroup*treatment interaction will also be presented.

7.7.5 Exploratory Analysis

Exploratory analysis will be performed on patients in the FITT set (including F1/F2/F3 patients) and the ITTF1 set unless otherwise specified.

For all ITTF1 summaries, fibrosis stage will not be included in the specified models.

Results collected at the end of the 72 week treatment period will be summarized.

7.7.5.1 Resolution of NASH without worsening of fibrosis

The number and percentage of patients with resolution of NASH without worsening of fibrosis will be summarized by treatment group.

Patients with missing data for resolution of NASH without worsening of fibrosis will be treated as non-responders for the main analysis.

The number of patients with resolution of NASH without worsening of fibrosis will be analysed using the same method as for the primary analysis of the primary endpoint (see section 7.7.1). The estimate of difference (elafibranor/placebo), its 95% CI and the corresponding p-value will be presented.

7.7.5.2 Improvement in fibrosis according to NASH CRN scoring

The number and percentage of patients with improvement in fibrosis according to NASH CRN scoring will be summarized by treatment group.

Patients with missing data for improvement in fibrosis according to NASH CRN scoring will be treated as non-responders for the main analysis.

The number of patients with an improvement in fibrosis according to NASH CRN scoring will be analysed using the same method as for the primary analysis of the primary endpoint (see section 7.7.1). The estimate of difference (elafibranor/placebo), its 95% CI and the corresponding p-value will be presented.

7.7.5.3 Other endpoints

All summaries within this section will be presented for the ITTF1 and FITT sets.

The number and percentage of patients with each of the following will be presented by treatment group at the end of the 72 week treatment period:

- Improvement of fibrosis of at least 1 stage according to NASH CRN scoring without worsening of NASH
- No worsening of fibrosis of at least 1 stage and no worsening of NASH
- Improvement of fibrosis of at least 1 stage according to NASH CRN scoring and resolution of NASH
- Improvement of fibrosis of at least 1 stage according to NAFLD Ishak scoring
- Improvement of NAS of at least 1 point
- Improvement in steatosis (NAS/SAF) of at least 1 grade
- Improvement of ballooning (NAS/SAF) of at least 1 grade
- Improvement of lobular inflammation (NAS) of at least 1 grade
- Improvement of lobular inflammation (SAF) of at least 1 grade
- Improvement in portal inflammation of at least 1 grade
- Improvement in NAS of at least 2 points
- Improvement of NAS of at least 2 points and with at least 1 point improvement in hepatic ballooning
- Improvement in NAS disease activity of at least 1 point
- Improvement in NAS disease activity score of at least a 1 point and improvement in hepatic ballooning of at least 1 point
- Improvement in NAS disease activity of at least 2 points
- Improvement in SAF activity score of at least 1 point
- Improvement in SAF activity score of at least a 1 point and improvement in hepatic ballooning of at least 1 point
- Improvement in SAF activity score of at least 2 points

Note: An improvement is defined as a reduction in stage/grade/points.

The number of patients with a response (e.g., resolution/improvement) for each endpoint will be analysed using the same method as for the primary analysis of the primary endpoint (see section 7.7.1). If appropriate, the endpoint specific baseline value will be introduced into the logistic model as covariate instead of the baseline NAS (categorical).

The estimate of difference (elafibranor/placebo), its 95% CI and the corresponding p-value will be presented. The response rate will also be presented for each treatment group with unadjusted 95% CIs using the Wilson score method.
Values for each of the following endpoints at baseline, at the end of the 72 week treatment period and change from baseline to end of 72 week treatment period, and shift tables of the number and percentage of patients with each value at baseline compared to the end of the 72 week treatment period will be summarized by treatment group as continuous data:

- Fibrosis according to NASH CRN scoring
- Fibrosis according to NAFLD Ishak scoring
- NAS
- NAS disease activity score
- SAF activity score
- Steatosis (NAS/SAF)

- Hepatic ballooning (NAS/SAF)
- Lobular inflammation (NAS)
- Lobular inflammation (SAF)
- Portal inflammation
- Area of fibrosis (normalized Collagen Proportional area in %) assessed by morphometry
- Framingham risk scores
- Liver stiffness assessed by FibroScan measurements

The change from baseline in each of the above endpoints at the end of the 72 week treatment period will be analysed using non-parametric ANCOVA method of Koch et al., with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F1, F2, F3), and baseline endpoint specific value (except for fibrosis where it will be baseline NAS, as specified in section 7.7.2). The statistical models will be used to calculate the treatment difference (elafibranor - placebo), 95% CI and p-value at each time-point. The treatment difference, 95% CI and corresponding p-values will be presented.

Values for each of the following endpoints at baseline, at each post-baseline time-point, and change from baseline to each post-baseline time-point will be summarized by treatment group as continuous data:

- Liver enzymes and liver markers
- Noninvasive markers of fibrosis and necrosis
- Lipid parameters
- Body weight
- Insulin resistance and glucose homeostasis markers
- Inflammatory markers
- SF-36 quality of life scores

The change from baseline in each endpoint at each time-point will be analysed using a repeated measures ANCOVA model with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), fibrosis stage (F1, F2, F3), endpoint specific baseline value, time-point and treatment by time-point interaction. Covariance parameters will be estimated via the REML method. Compound symmetry will be used for the covariance structure. The Kenward and Roger's method for evaluating the degrees of freedom will also be used.

The statistical model will be used to calculate the treatment difference (elafibranor - placebo), 95% CI and p-value at the end of the 72 week treatment period. The LSMean and 95% CI for each treatment group, LSMean treatment difference, 95% CI and corresponding p-values will also be presented at each time-point.

Note: Please see section 7.7.1 and 7.7.2 for further details.

7.8 Safety

Unless otherwise stated, all safety summaries will be presented for patients in the Safety, FSS and SSF1 populations.

7.8.1 Extent of Exposure

Duration of exposure will be defined in weeks as:

([last dose date – first dose date] + 1– off-drug days) / 7

If date of first dose date is missing, then date of randomization visit will be used. If last dose date is missing, then date of last known dose will be used.

Duration of exposure will be listed and summarized using descriptive statistics for each treatment group, and overall.

The number and percentage of patients with duration of exposure in the following categories will be summarized:

- ≥0 and <12 weeks
- ≥12 and <24 weeks
- ≥24 and <36 weeks
- ≥36 and <48 weeks
- ≥48 and <60 weeks
- ≥60 and <72 weeks
- ≥72 and <96 weeks
- ≥96 and <120 weeks
- ≥120 and <144 weeks
- ≥144 and <168 weeks
- ≥168 and <192 weeks
- ≥192 and <216 weeks
- ≥216 and <240 weeks
- ≥240 and <264 weeks
- ≥264 and <288 weeks
- ≥288 and <312 weeks
- ≥312 weeks

The extent of exposure will also be displayed in the subset of patients from the ITT set with missing biopsy results at Week 72.

7.8.2 Adverse Events

All adverse events (AEs) recorded on the eCRF will be coded using the MedDRA dictionary [Version 19.0 (or a later version if updated during the study)] and classified as either non treatment-emergent AEs (pre-treatment AEs, post-treatment AEs) or treatment emergent AEs (TEAEs) as follows:

- Pre-treatment AEs are events that start prior to the date of first dose of study treatment.
- Post-treatment AEs are events that start more than 30 days after stop of study treatment.

- TEAEs are events with start date on or after the date of first dose of study treatment and up to 30 days after date of last dose of study treatment or events with start date prior to the date of first dose of study treatment whose severity worsens on or after the date of first dose of study treatment.

In regards to the change in definition of the long-term primary endpoint as documented in the protocol amendment 4, the following events will not have to be considered as efficacy endpoints but as AEs included in safety analyses in addition to the AEs reported by the investigators:

- hepatocellular carcinoma;
- onset of hepatorenal syndrome;
- onset of hepatopulmonary syndrome;
- onset of chronic gastrointestinal blood loss due to portal hypertensive gastropathy and leading to hospitalization or transfusion.

In case of above events recorded as clinical outcome and not as AEs at the time of the SEA, AE records will be constructed based upon known case reporting to reflect the event. In case of missing detailed information (e.g. seriousness, intensity …), conservative clinical interpretation will be utilized in order to accurately define the fields for these records. Further information regarding this process will be documented in a dedicated data transfer agreement.

All AE data will be listed by treatment group. Treatment-emergence status will be flagged in the listing. In addition, corresponding listings of serious AEs (SAEs), AEs leading to discontinuation of study treatment and AEs resulting in death will be produced.

The number and percentage of patients reporting each pre-treatment AE will be summarized for each treatment group and overall, by SOC (sorted alphabetically) and PT (sorted by descending overall total).

Summary tables of TEAEs by treatment group and overall will be produced.

The severity of all AEs is recorded as mild, moderate, or severe. If severity is missing for a TEAE, it will be considered severe only in the overall category in the summary tables.

The relationship between an AE and treatment is assessed as 'not related, unlikely related, possibly related, related, and not assessable'. A treatment-related AE is an AE considered by the investigator as possibly related, related to treatment, and not assessable, or with unknown/missing relationship to treatment.

An overview table will summarize the number and percentage of patients with at least one of the following TEAEs, where patients with more than one TEAE in a particular category are counted only once in that category:

- any AE;
- any AE by severity (mild, moderate, severe);
- treatment-related AE;
- AE leading to treatment discontinuation;
- treatment related AE leading to discontinuation;
- SAE;
- treatment-related SAE;

- SAE leading to death;
- treatment-related SAE leading to death;
- SAE leading to treatment discontinuation;

The overview table will also summarize the number of the following TEAEs:

- all AEs;
- AEs by severity (mild, moderate, severe);
- treatment-related AEs;
- treatment-related AEs by severity (mild, moderate, severe);
- SAEs;
- SAEs by severity (mild, moderate, severe);
- treatment-related SAEs.

The number and percentage of patients reporting each TEAE and the total count of TEAEs will be summarized by SOC and PT. Tables will be sorted alphabetically by SOC. PTs will be sorted by descending frequency overall total. The following summaries will be produced:

- AEs by SOC and PT;
- Most common AEs (≥ 5% in any treatment group) by SOC and PT;
- AEs related to treatment, by SOC and PT;
- AEs by relationship to treatment, by SOC and PT;
- AEs by maximum severity, by SOC and PT;
- AEs related to treatment by maximum severity, by SOC and PT;
- AEs causing discontinuation from treatment, by SOC and PT;
- AEs related to study treatment causing discontinuation from treatment, by SOC and PT;
- AEs causing treatment delays and interruptions, by SOC and PT;
- SAEs, by SOC and PT;
- SAEs related to treatment, by SOC and PT;
- AEs leading to death, by SOC and PT.

In the above summaries, patients with more than one AE within a particular SOC are counted only once for that SOC. Similarly, patients with more than one AE within a particular PT are counted only once for that PT. For summaries by maximum severity, patients with multiple AEs within a particular SOC or PT will be counted under the category of their most severe AE within that SOC or PT. AEs with missing intensity/severity will be included (as severe) in the overall count of patients with AEs.

AEs of special interest (AESI) are TEAEs which are defined according to categories and sub-categories as follows:

- CPK elevations of severe intensity or leading to permanent study drug discontinuation
- Muscle injury symptoms of severe intensity corresponding to:
  - Muscle pain or Myalgia
  - Muscle spasms or Tremor
  - Muscle weakness
- Transaminases elevations from baseline of severe intensity or leading to permanent study drug discontinuation

- Liver injury events of severe intensity corresponding to:
  - Hepatic impairment
  - Hepatic failure
- Gastrointestinal symptoms of severe intensity corresponding to:
  - Abdominal pain
  - Constipation
  - Diarrhea
  - Nausea
  - Vomiting
  - Acute cholecystis
  - Acute pancreatitis
- Fatigue and Asthenia of severe intensity
- Serum creatinine elevations of severe intensity or leading to permanent study drug discontinuation
- Renal injury events of moderate or severe intensity corresponding to:
  - Renal failure
  - Renal impairment
  - Renal colic
- Treatment emergent Pregnancy and outcomes of Pregnancy

All of the above categories, with the exception of the pregnancy AESIs, will be identified via a list of MedDRA queries of PT codes. Pregnancy AESIs will be identified using positive pregnancy assessment and reported in the pregnancy report form.

For each of the AESI categories and subcategories defined above, besides treatment emergent Pregnancy, the number and percentage of patients with at least one event and the corresponding number of events will be summarized for any:

- AESIs;
- Treatment related AESIs;
- Serious AESIs;
- AESIs leading to death;
- AESIs leading to treatment discontinuation.

In addition, the total number and percentage of patients with at least one AESI (with the exception of treatment emergent Pregnancy), and the corresponding number of events will be summarized for the five categories above.

For Pregnancy, the number and percentage of patients with at least one event and the corresponding number of events will be summarized. The outcomes of Pregnancy will also be summarized by SOC and PT.

In addition, the following summaries will be produced for the post-treatment AEs:

- Overall summary of post-treatment AEs;
- Overall summary of number of post-treatment AEs;
- AEs by SOC and PT;
- SAEs by SOC and PT.

No statistical comparisons of AEs between treatment groups will be performed.

The following summaries will also be displayed in the subset of patients from the ITT set with missing biopsy results at Week 72:

- The overview summarizing the number and percentage of patients with at least one TEAE;
- The overview summarizing the number and percentage of patients with at least one AESI for each category and subcategory;
- The overview of the number of TEAEs;
- The overview of the number of AESIs for each category and subcategory;
- The number and percentage of patients reporting at least one TEAE and the corresponding count of TEAEs summarized by SOC and PT.

7.8.3 Laboratory Evaluations

Clinical laboratory data to be summarized includes hematology, blood chemistry, and urinalysis parameters received from the central laboratory.

The table below presents the clinical laboratory tests to be performed as specified in the protocol.

Table 2 Clinical Laboratory Tests

Clinical Laboratory Test	Parameter
Hematology	Differential Count Hemoglobin Hematocrit Platelet Count PT (INR) Red Blood Cell Count Reticulocytes White Blood Cell Count
Serum Chemistry	Alanine aminotransferase (ALT) Albumin Alkaline phosphatase Apolipopprotein (ApoAI and ApoB) Aspartate aminotransferase (AST) Blood Urea Nitrogen Calcium Chloride Creatinine Creatine phosphokinase Direct Bilirubin (Conjugated Bilirubin in the protocol) Estimated glomerular filtration rate [MDRD, CKD-EPI (Creatinine), CKD-EPI (Cystatin C)][a] Fasting plasma glucose Gamma-Glutamyltransferase (GGT)

	Glucose Insulin Hemoglobin A1c High-sensitivity C-reactive protein High density lipoprotein-C Homeostasis model assessment of insulin resistance Low density lipoprotein-C Non High density lipoprotein-C Triglycerides Total cholesterol Total Bilirubin Very low density lipoprotein-C
Urinalysis	Albumin Albumin-creatinine ratio Creatinine Interleukin-18 Kidney injury molecule-1 Microscopic analysis a1 microgobulin Neutrophil gelatinase-associated lipocalin (NGAL) N-Acetyl-β-glucosaminidase (beta-NAG)
Urinalysis (dipstick)	Bilirubin Blood Glucose Ketones Leukocytes Nitrite pH Protein Specific gravity Urobilinogen
[a] Estimate glomerular filtration rate (mL/min/1.73m^2) is derived as: 133 x [minimum of Standardized serum cystatin C [mg/l] / 0.8, 1]$^{-0.499}$ x [maximum of Standardized serum cystatin C [mg/l] / 0.8, 1]$^{-1.328}$ x 0.996$^{Age\ (years)}$ x 0.932 [if female]	

All laboratory data will be reported in International System of Units (SI) units.

All laboratory data will be listed. Out-of-reference-range values will be flagged as high (H) or low (L) in the listings.

For analysis purposes, values preceded by a "<" or a ">" sign (i.e. those below or above the limits of quantification) will be considered equal to the lower or upper limit of quantification, respectively.

Only clinical laboratory tests included in the table will be included in the summary tables.

Clinical laboratory values recorded at each time-point (baseline, week 24, week 48, week 72, week 96, week 120, week 144, week 168, week 192, week 216, week 240, week 264, week 288, week 312 and at the

end of the 72 week treatment period) and change from baseline to each post baseline time-point will be summarized by treatment group and overall as continuous data.

Line plots of means with whiskers extending to the mean +/ SE over time using the same visits as for table summaries will be presented for hemoglobin, hematocrit, RBC count, ALT, AST, GGT, alkaline phosphatase, creatinine, non high density lipoprotein-C, high density lipoprotein-C, low density lipoprotein-C, very low density lipoprotein-C, triglycerides, total cholesterol, a1 microgobulin, beta-NAG, NGAL, urinary creatinine and urinary albumin-creatinine ratio.

For hematology and serum chemistry, shift tables presenting movement in and out of reference range from baseline to end of the 72 week treatment period and to the last post-baseline visit will be provided for each treatment group. Corresponding shift tables normal vs. abnormal will be produced for quantitative urinalysis parameters.

eDISH plots will present the concomitant values of aminotransferase (ALT and AST) and total bilirubin at the visit corresponding to the peak of aminotransferase. Distinct plots will be produced separately with regards to the peaks of ALT and AST. For each subject and for both ALT and AST, the strategy is as follows:

1) Among all the visits, identify the maximum value of aminotransferase
2) If the peak value occurs at just one visit, keep the value of total bilirubin at the visit corresponding to the peak of aminotransferase
3) If the peak value occurs at more than one visit, keep the highest value of total bilirubin at any of the visits corresponding to the peak of aminotransferase
4) If the peak value occurs at a visit / visits with no corresponding value/values of total bilirubin, keep the value of bilirubin closest to the corresponding visit / any of the corresponding visits. In the event of equidistant time from the peak value visit / visits, keep the highest value of total bilirubin at any of the visits equidistant from the corresponding visit/visits of peak aminotransferase
5) Classify subjects into two subgroups: baseline value of ALT < ULN or baseline value of ALT ≥ ULN.
   a. If baseline value of ALT < ULN, express total bilirubin in xULN on the y-axis and aminotransferase in xULN on the x-axis. Add horizontal and vertical lines corresponding to total bilirubin = 2 ULN and aminotransferase = 3 ULN respectively.
   b. If baseline value of ALT ≥ ULN, present two plots:
      i. Express total bilirubin in xBaseline on the y-axis and aminotransferase in xBaseline on the x-axis. Add horizontal and vertical lines corresponding to total bilirubin = 2xBaseline and aminotransferase = 3xBaseline respectively. For total bilirubin values > 2xBaseline, values identified as > 2 ULN will be plotted differently than values ≤ 2 ULN.
      ii. Express total bilirubin in xULN on the y-axis and aminotransferase in xBaseline on the x-axis. Add horizontal and vertical lines corresponding to total bilirubin = 2 ULN and aminotransferase = 3xBaseline respectively.

Occurrence of aminotransferases (ALT and AST) increase will be presented separately according to the baseline level (normal or abnormal). The classes considered aminotransferase increase are:

1) > 3ULN and ≤ 5ULN, > 5ULN and ≤ 10ULN, > 10ULN if the baseline level of aminotransferase is normal (i.e. < ULN). Among all visits, the worst case will be retained.

2) > 3xBaseline and ≤ 5xBaseline, > 5xBaseline if the baseline level of aminotransferase is abnormal (i.e. ≥ ULN). Among all visits, the worst case will be retained.

Occurrence of CPK increase will be presented according to the baseline level (normal or abnormal). The classes considered for CPK increase are: > 3ULN and ⩽ 5ULN, > 5ULN. Among all visits, the worst case will be retained.

The increase in serum creatinine will be presented as:

1) Occurrence of at least one post baseline value > ULN (only on the subset of subjects with baseline value ≤ ULN)
2) Change from baseline in creatinine presented in classes according to KDIGO (Kidney Disease Improving Global Outcomes) AKI (Acute Kidney Injury) stages. Among all visits, the worst case will be retained:
    a. ≥ 1.5xBaseline and < 2.0xBaseline, or ≥ 0.3 mg/dL increase
    b. ≥ 2.0xBaseline and < 3.0xBaseline
    c. ≥ 3.0xBaseline or ≥ 4 mg/dL.

The relative decrease from baseline in eGFR (estimated glomerular filtration rate) will be presented in classes: ≥ 0 % and < 25 %, ≥ 25 % and < 50 %, ≥ 50 % and < 75 %, ≥ 75 %. Among all visits, the worst case will be retained.

In addition, the number and percentage of patients with markedly abnormal clinical laboratory values will be summarized for each parameter by time-point.

For each laboratory analyte, the baseline value will be defined as last scheduled or unscheduled value collected prior to the first dose of study treatment. Assessments carried out on day of first study treatment administration are considered to have taken place before the study treatment administration, if the corresponding times have not been recorded.

7.8.4 Vital Signs

Vital Signs to be summarized include:

- systolic blood pressure (mmHg);
- diastolic blood pressure (mmHg);
- pulse rate (bpm);
- weight (kg);
- waist circumference (cm);

Vital sign data recorded at each time-point (baseline, week 12, week 24, week 36, week 48, week 60, week 72, week 96, week 120, week 144, week 168, week 192, week 216, week 240, week 264, week 288, week 312, at the end of the 72 week treatment period) and change from baseline to each post baseline time-point will be summarized by treatment group and overall as continuous data.

Note: waist circumference is not collected at week 12, week 36 or week 60, so will not be included in the summary tables at these time-points.

The baseline value will be defined as last scheduled or unscheduled value collected prior to the first dose of study treatment. Assessments carried out on day of first study treatment administration are considered to have taken place before the study treatment administration, if the corresponding times have not been recorded.

7.8.5 Electrocardiograms

An overall Investigator assessment of ECG will be provided (categories "normal", "abnormal, not clinically significant" and "abnormal, clinically significant").

The Investigator assessment will be listed and the number and percentage of patients within each assessment category will be tabulated by treatment group and overall and visit (baseline, week 36, week 72, week 120, week 168, week 216, week 264, week 312 and at the end of the 72 week treatment period).

Shifts from baseline (normal vs. abnormal, not clinically significant vs. abnormal, clinically significant) at the end of the 72 week treatment period and to the last post-baseline visit will be presented.

7.8.6 Physical Examination

Abnormalities identified from physical examination are recorded in the eCRF as Medical History or Adverse Events as appropriate and will be listed and summarized as such [See Sections 7.5.1 (Medical History) and 7.8.2 (Adverse Events)].

Physical examination results (normal/abnormal) and details of abnormalities will be listed for each subject.

7.8.7 Other Safety Variables

7.8.7.1 Other Safety Markers

Other safety markers to be summarized include NT-ProBNP, troponin-T, Homocysteine and cystatin C. NT-ProBNP will be log-transformed for the purpose of summaries and analysis.

Other safety marker values recorded at each time-point (baseline, week 24, week 48, week 60, week 72, week 96, week 120, week 144, week 168, week 192, week 216, week 240, week 264, week 288, week 312 and at the end of the 72 week treatment period) and change from baseline to each post baseline time-point will be summarized by treatment group and overall, and by endpoint, as continuous data.

Line plots of means with whiskers extending to the mean +/ SE over time using the same visits as for table summaries will be presented for homocysteine and cystatin C.

7.8.7.2 Drug Induced Liver Injury

As stated in the protocol, the DILI will be adjudicated by the Clinical Events Committee (CEC). The number and percentage of patients with adjudicated DILI events confirmed by the CEC will be summarized by treatment group.

7.9 Interim Analysis

For details on the Interim Analysis, please see the Final Analysis Statistical Analysis Plan.

8 CHANGES IN PLANNED ANALYSES

The exploratory F1 cohort analysis set defined in the protocol has been removed and replaced with the ITTF1 cohort set and additionally the SSF1 has also been added.

9 DATA ISSUES

There are no data issues.

10 REFERENCES

- D'Agostino RB Sr, Vasan RS, Pencina MJ, et. al. General Cardiovascular Risk Profile for Use in Primary Care. The Framingham Heart Study. *Circulation*. 2008 Jan 22.
- ICH. *Statistical Principles for Clinical Trials*, Guideline E9, 1998. Available at http://www.emea.eu.int/pdfs/human/ich/036396en.pdf
- ICH E9 (R1). *Addendum to Statistical Principles for Clinical Trials on Choosing Appropriate Estimands and Defining Sensitivity Analyses in Clinical Trials*, 2017. Available at https://www.ema.europa.eu/en/documents/scientific-guideline/draft-ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical_en.pdf
- Dmitrienko, A., Tamhane, A.C. Mixtures of multiple testing procedures for gatekeeping applications in clinical trials. *Statistics in Medicine.* 2011 30: 1473-1488.
- Dmitrienko, A., Tamhane, A.C. General theory of mixture procedures for gatekeeping. *Biometrical Journal.* 2013 55: 402-419.
- CPMP. *Points to Consider on Missing Data*. EMEA: London, 2001. Available at http://www.emea.eu.int/pdfs/human/ewp/177699EN.pdf
- CPMP. *Points to Consider on Adjustment for Baseline Covariates*. EMEA: London, 2003
- ICH. *ICH E3 Guideline: Structure and Content of Clinical Study Reports Questions & Answers*, 2012. Available at http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E3/E3_QAs_R1_Step4.pdf
- KDIGO. *Kidney Disease: Improving Global Outcomes (KDIGO) clinical practice guideline for acute kidney injury. Kidney International Supplement*. 2012; 2: 19-36. Available at https://kdigo.org/wp-content/uploads/2016/10/KDIGO-2012-AKI-Guideline-English.pdf
- Ge, M., Durham, L. K., Meyer, R. D., Xie, W., & Thomas, N. (2011). Covariate-adjusted difference in proportions from clinical trials using logistic regression and weighted risk differences. *Drug Information Journal*, *45*(4), 481-493.

- Stokes, Maura E., Charles S. Davis, and Gary G. Koch. *Categorical data analysis using SAS*. SAS institute, 2012
- Dmitrienko A and Koch G. *Analysis of Clinical Trials using SAS A Practical Guide 2nd ed.*, SAS Institute July 2017

11 APPENDICES

Appendix I - Schedule of Events

Table 3: STUDY GENERAL ASSESSMENT SCHEDULE

	Screening Period			First Treatment Period							Long-term Treatment Period		End Of Study Treatment
Visit	**SV1**	**SV2** [4]	**SPV** [5]	**V1**	**V2**	**V3**	**V4**	**V5**	**V6**	**V7**	**Phone Visit PV1-PVn**	**V8-Vn**	**EOT** [13]
Week	*-12 to -8*	*-12 to -4*	*-1* [5]	*0* [6]	*12* [6]	*24* [6]	*36* [6]	*48* [6]	*60* [6]	*72* [6]	*(every 24 weeks starting 12 weeks after V7)* [8]	*(every 24 weeks starting after V7)* [9]	*30 days after final study drug administration*
Permitted Margin				0	±1 week after V1	± 1 week after V1	± 1 week after V1	± 1 week after V1	± 1 week after V1	± 1 week after V1	± 2 weeks Compared to V7	± 2 weeks Compared to V7	± 1 week after last administration
Obtain informed consent	X												
Medical history / demographics	X												
Check inclusion / exclusion criteria	X			X [7]									
Adequate diet and lifestyle recommendations, including alcohol restrictions and smoking habits	X - - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - - →	
Confirmation of diet and lifestyle compliance, including alcohol restrictions and smoking habits				X	X	X	X	X	X	X	X	X	X
Physical examination	X			X	X	X	X	X	X	X		X	X
Vital signs & height [1] & weight measurement	X			X	X	X	X	X	X	X		X	X
Waist circumference	X			X		X		X		X		X	X

	Screening Period			First Treatment Period							Long-term Treatment Period		End Of Study Treatment
Visit	**SV1**	**SV2** [4]	**SPV** 5	**V1**	**V2**	**V3**	**V4**	**V5**	**V6**	**V7**	**Phone Visit PV1-PVn**	**V8-Vn**	**EOT** [13]
Week	*-12 to -8*	*-12 to -4*	*-1* **5**	*0* **6**	*12* **6**	*24* **6**	*36* **6**	*48* **6**	*60* **6**	*72* **6**	*(every 24 weeks starting 12 weeks after V7)* **8**	*(every 24 weeks starting after V7)* **9**	*30 days after final study drug administration*
Permitted Margin				**0**	**±1 week after V1**	**± 1 week after V1**	**± 1 week after V1**	**± 1 week after V1**	**± 1 week after V1**	**± 1 week after V1**	**± 2 weeks Compared to V7**	**± 2 weeks Compared to V7**	**± 1 week after last administration**
12-Lead ECG				**X**			**X**			**X**		**X** [10]	**X**
Lab evaluation (see Table 4)	**X**	**X**		**X**	**X**	**X**	**X**	**X**	**X**	**X**		**X**	**X**
Send sample for central histological evaluation of NASH diagnosis / change	**X**	**X**								**X**		**X** [11]	
Liver biopsy		**X** [4]								**X**		**X** [11]	
Phone call to patient to confirm eligibility of histology criteria			**X** 5										
FibroScan [2]				**X**						**X**		**X**	
Contact the patient prior to visit [3]				**X**	**X**	**X**	**X**	**X**	**X**	**X**		**X**	**X**
Randomization				**X**									
IXRS registration	**X**			**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
Review prior / concomitant medication	**X**			**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**

	Screening Period			First Treatment Period							Long-term Treatment Period		End Of Study Treatment
Visit	**SV1**	**SV2** [4]	**SPV 5**	**V1**	**V2**	**V3**	**V4**	**V5**	**V6**	**V7**	**Phone Visit PV1-PVn**	**V8-Vn**	**EOT** [13]
Week	*-12 to -8*	*-12 to -4*	*-1* **5**	*0* **6**	*12* **6**	*24* **6**	*36* **6**	*48* **6**	*60* **6**	*72* **6**	*(every 24 weeks starting 12 weeks after V7)* **8**	*(every 24 weeks starting after V7)* **9**	*30 days after final study drug administration*
Permitted Margin				**0**	**±1 week after V1**	**± 1 week after V1**	**± 1 week after V1**	**± 1 week after V1**	**± 1 week after V1**	**± 1 week after V1**	**± 2 weeks Compared to V7**	**± 2 weeks Compared to V7**	**± 1 week after last administration**
Quality of life assessment				X		X		X		X		X [12]	X
Adverse events	X	X		X	X	X	X	X	X	X	X	X	X
Data collection on clinical outcomes					X	X	X	X	X	X	X	X	X
Study placebo or drug dispensation				X	X	X	X	X	X	X		X	
Drug accountability					X	X	X	X	X	X	X	X	X

Abbreviations: ECG = electrocardiogram; EOT = end of treatment; IXRS = Interactive voice/web Response System; LTTP = Long-term Treatment Period; NASH = nonalcoholic steatohepatitis; PV = phone visit; QOL = quality of life; SV = Screening visit; V = visit

1. Height is measured only at visit SV1.

2. Where possible FibroScan must be done at the day of visit. Otherwise, it can be performed within 7 days around the visit date.

3. During the study, the patient should be contacted at least 1 week before the next visit as a reminder on procedures and IP return.

4. This visit only occurs if no historical biopsy within 6 months before the Screening Visit is available. A screening liver biopsy and slides shipment to the central anatomopathologist must be performed at least 4 weeks before Randomization (in order to obtain the results in time). Coagulation (platelet count and PT [INR]) should be checked locally prior to this liver biopsy (according to local medical standards in each hospital).

5. Screening Phone Visit. Telephone contact for all patients at least 1 week before V1. Patients should be contacted regarding eligibility confirmation within 1 week prior to Randomization Visit V1. In case of ineligibility, the patient should be contacted as soon as possible.

6. The maximum time period between visits in the First Treatment Period is to be 96 days due to the study drug supply provided to the patient.

7. Check of all inclusion/exclusion criteria, including biological and histological criteria assessed at SV1 and SV2.

8. Phone visits every 24 weeks starting 12 weeks after V7 for safety, data collection on clinical outcomes, and IP compliance control. If any information during this phone contact requires a visit, the Investigator may decide to perform an unscheduled visit. Phone visits may also be performed at the same frequency for the follow-up of patients having permanently discontinued study drug but remaining in the study (Same information collected except IP compliance control).

9. The maximum time period between visits in the Long-term Treatment Period (LTTP) is to be 192 days due to the study drug supply provided to the patient.

10. In the LTTP the first ECG will be performed at V9 and then every 48.

11. Liver biopsy will be performed after approximately 4 years of treatment (V13) and in the case of suspicion of cirrhosis (based on FibroScan and/or clinical or biological assessment). In the case that the suspicion of cirrhosis is not confirmed by liver biopsy the patient shall remain on the study and a liver biopsy will be performed after approximately 4 years of treatment (V13, unless a biopsy has already been performed within the year). Blood sampling (coagulation tests; see Table 4) are to be performed locally before the biopsy.

12. QOL assessment questionnaire to be completed at 24 (V8), 48 (V9), and 96 (V11) weeks in the LTTP (following approximately 96, 120, and 168 weeks of treatment, respectively), and every 48 weeks thereafter.

13. EOT Visit to be performed 30 days after final study drug administration at the end of study or for any premature discontinuation (permanent study drug discontinuation or trial discontinuation).

Table 4: STUDY BIOLOGICAL ASSESSMENT SCHEDULE

Visit	Screening Period		First Treatment Period							Long-term Treatment Period	End of Study Treatment
	SV1 SB1 6	SB2 7	V1 B1	V2 B2	V3 B3	V4 B4	V5 B5	V6 B6	V7 B7	V8 to Vn B8 to Bn	EOT
Week	*-12 to -8*	*Prior to -4*	*0*	*12*	*24*	*36*	*48*	*60*	*72*	*Every 24 weeks*	*30 days after final study drug administration*
Hematology *Hemoglobin, hematocrit, RBC, WBC, differential count, platelet count, reticulocytes count, and PT (INR)*	X		X	X	X	X	X	X	X	X	X
Coagulation - local lab testing prior to liver biopsy *Platelet count, PT (INR)* 1		X							X	X 1	
Serology *HIV ab I/II, HBsAg, and HCV Ab (positive HCV RNA in case HCV Ab >0 or known cured hepatitis C infection* 2*)*	X										
Screening Visit 1 - chemistry panel *HbA1c*2*, fasting plasma glucose, insulin (fasting), HOMA-IR creatinine, eGFR, GGT, AST, ALT, CPK*2*, alkaline phosphatase, total and conjugated bilirubin, sodium, TG, and MELD score*	X										
V1 to Vn total chemistry panel *HbA1c, fasting plasma glucose, creatinine, eGFR, GGT, AST, ALT, CPK, alkaline phosphatase, total proteins, albumin, electrolytes (sodium, potassium, chloride, calcium), uric acid, urea (BUN), total and conjugated bilirubin, hsCRP, total cholesterol, nonHDL-C, HDL-C, TG, calculated VLDL-C, ApoAI, ApoB, calculated LDL-C, and MELD score*			X 6	X	X	X	X	X	X	X	X

Visit	Screening Period		First Treatment Period							Long-term Treatment Period	End of Study Treatment
	SV1 SB1 6	SB2 7	V1 B1	V2 B2	V3 B3	V4 B4	V5 B5	V6 B6	V7 B7	V8 to Vn B8 to Bn	EOT
Week	*-12 to -8*	*Prior to -4*	*0*	*12*	*24*	*36*	*48*	*60*	*72*	*Every 24 weeks*	*30 days after final study drug administration*
Urinalysis *albumin, creatinine, ACR, and microscopic analysis* *α1 microglobulin*, β-NAG,* N-Gal*, IL-18*, KIM-1*,*			X	X	X	X	X	X	X	X	X
Urinalysis (dipstick) *Specific gravity, pH, protein, glucose, ketones, bilirubin, urobilinogen, blood, nitrite, and leukocytes*	X		X	X	X	X	X	X	X	X	X
Urinary pregnancy tests 3	X		X	X	X	X	X	X	X	X	X
Inflammatory markers *Fibrinogen, and haptoglobin*			X		X		X		X	X	X
Other Liver markers *CK18 (M65 & M30), adiponectin, ferritin, FGF19 & FGF21, alpha2 macroglobulin, hyaluronic acid, PIIINP, TIMP-1, and CHI3L1* 4			* 4		*		*		* 4	* 4	*
Calculated fibrosis & steatosis index *Fibrotest, ELF, NAFLD Fibrosis score, Steatotest, FLI, Fibrometre S, and FIB-4*			*		*		*		*	*	*
Other safety markers *Homocysteine, NT-ProBNP, troponin-T, and cystatin C*			*		*		*		*	*	*
Special glycemic and other lipid parameters *Insulin(fasting), HOMA-IR, Fructosamine, C-peptide, FFA, small dense LDL, ApoAII, Apo CIII, and Apo E*			*		*		*		*	*	*

Visit	Screening Period		First Treatment Period							Long-term Treatment Period	End of Study Treatment
	SV1 SB1 6	SB2 7	V1 B1	V2 B2	V3 B3	V4 B4	V5 B5	V6 B6	V7 B7	V8 to Vn B8 to Bn	EOT
Week	*-12 to -8*	*Prior to -4*	*0*	*12*	*24*	*36*	*48*	*60*	*72*	*Every 24 weeks*	*30 days after final study drug administration*
Sampling for additional parameters *Whole blood 5, plasma, and serum bank*	* 5		*	*	*	*	*	*	*	*	*

X = results available within 2 working days (routine analysis) * = batch analysis

Abbreviations Ab = antibody; ACR = albumin–creatinine ratio; Ag = antigen; ALT = alanine aminotransferase; Apo = apolipoprotein; AST = aspartate aminotransferase; β-NAG = N-acetyl-β-D-glucosaminidase; BUN = blood urea nitrogen; B = biological assessment Visit; CHI3L1 = chitinase-3-like protein 1; CK18 = cytokeratin 18; CPK = creatine phosphokinase; eGFR = estimated glomerular filtration rate; ELF = enhanced liver fibrosis; EOT = end of study treatment; FFA = free fatty acid; FGF = fibroblast growth factor; FIB-4 = fibrosis 4 score; FLI = fatty liver index; GGT = gamma-glutamyl transferase; HbA1c = hemoglobin A1c; HBsAg = hepatitis B surface antigen; HCV = hepatitis C virus; HDL-C = high density lipoprotein-C; HIV = human immunodeficiency virus; HOMA-IR = homeostasis model assessment of insulin resistance; hsCRP = high-sensitivity C-reactive protein; IL-18 = interleukin 18; INR = international normalized ratio; KIM-1 = kidney injury molecule-1; LDL-c = low density lipoprotein-C; MDRD = modification of diet in renal disease; MELD = model end stage liver disease; NAFLD = nonalcoholic fatty liver disease; N-Gal = neutrophil gelatinase-associated lipocalin; NT-ProBNP = N-terminal of the prohormone brain natriuretic peptide; PIIINP = type III procollagen peptide; PT = prothrombin time; TIMP-1 = tissue inhibitors of metalloproteinases 1; RBC = red blood cell; SB = Screening biological assessment Visit; SV = Screening Visit; TG = triglyceride; VLDL-C = very low density lipoprotein-C; V = Visit; WBC = white blood cell.

1. Coagulation (platelet count and PT [INR]) should be checked prior to any liver biopsy (according to local medical standards in each hospital). To be done through a local laboratory. Liver biopsy will be performed after 72 weeks (V7) and after approximately 4 years of treatment (V13) and in the case of suspicion of cirrhosis (based on FibroScan and/or clinical or biological assessment). In the case that the suspicion of cirrhosis is not confirmed by liver biopsy the patient shall remain on the study and a liver biopsy will be performed after approximately 4 years of treatment ([V13] unless a biopsy has already been performed within the past year).

2. Upon receipt of the results of the biological assessment performed at SV1, retesting or additional testing may be needed during the Screening Period:

   - CPK can be repeated prior to Randomization (V1) within 1 to 2 weeks after initial test.
   - HbA1c can be repeated prior to Randomization (V1), at the latest 2 weeks prior to planned Randomization.
   - HCV RNA can be tested, at SV1 in case of known cured hepatitis c infection, or in case of positive HCV Ab at SV1, at a retest screening visit at the latest 2 weeks prior to the planned Randomization (V1).

3. Dipstick at site for WOCBP only. In addition, home pregnancy tests are to be performed by WOCBP every 4 weeks from V1.

4. CHI3L1 to be tested only at V1, V7, and at the time of 4 years biopsy (V13).

5. Whole blood sample will be only taken at SV1 while plasma and serum samples are to be taken at every visit **ONLY** for patients who have signed the pharmacogenomic and biomarker ICF.

6. Visits SV1 and V1 should be scheduled at least 8 weeks apart in order to have 2 consecutive baseline values of AST, ALT, total bilirubin, and INR for DILI adjudication.

7. SB2, additional visit in the Screening Period if required for coagulation prior to liver biopsy.

Appendix II - Table, Figure and Listing Shells

The table, figure and listing shells and corresponding Table of Contents are in a separate file.

Appendix III – SF-36 Quality of Life Scoring Manual

The SF-36 quality of life scoring manual is in a separate file.

[redacted]

GENFIT

Protocol No.: GFT505-315-1

A Multicenter, Randomized, Double-Blind, Placebo-Controlled Phase III Study to Evaluate the Efficacy and Safety of Elafibranor in Patients with Nonalcoholic Steatohepatitis (NASH) and fibrosis

[redacted] Study ID: RESOLVE-IT/8325398

STATISTICAL ANALYSIS PLAN

LONG-TERM ENDPOINT ANALYSIS (FINAL ANALYSIS)

Version: V5.0
Date of Issue: 23-NOV-2020

Author: [redacted]

Genfit
Parc Eurasanté
885, Avenue Eugène Avinée
59120 LOOS – France

[redacted]

APPROVALS

The undersigned agree that all required reviews of this document are complete, and approve this Statistical Analysis Plan (SAP) as final. Programming of the tables, figures and listings based upon the specifications within this document can proceed.

[redacted] **Approval:**

[redacted] Dat[redacted]

Printed Name/Title
[redacted]

Genfit Approval:

Signature Date

Printed Name/Title
[redacted]

Signature Date

Printed Name/Title
[redacted]

APPROVALS

The undersigned agree that all required reviews of this document are complete, and approve this Statistical Analysis Plan (SAP) as final. Programming of the tables, figures and listings based upon the specifications within this document can proceed.

[redacted] **Approval:**

Signature Date

Printed Name/Title
[redacted]

Genfit Approval:

Signature Date
[redacted]

Signature Date
[redacted]

Printed Name/Title
[redacted]

REVIEWERS

The following reviews of the SAP were conducted:

Name and Title	Role	Version Last Reviewed	Company/ Organization
[redacted]	[redacted]	0.1	[redacted]
[redacted]	[redacted]	0.1	[redacted]
[redacted]	[redacted]	0.1	[redacted]
[redacted]	[redacted]	0.1	[redacted]
[redacted]	[redacted]	0.2	Genfit
[redacted]	[redacted]	0.2	Genfit
[redacted]	[redacted]	0.2	Genfit
[redacted]	[redacted]	0.2	[redacted]
[redacted]	[redacted]	0.3	[redacted]
[redacted]	[redacted]	0.3	Genfit
[redacted]	[redacted]	0.3	Genfit
[redacted]	[redacted]	0.3	[redacted]
[redacted]	[redacted]	0.3	[redacted]
[redacted]	[redacted]	0.4	[redacted]
[redacted]	[redacted]	0.4	Genfit
[redacted]	[redacted]	1.1	Genfit
[redacted]	[redacted]	1.1	Genfit
[redacted]	[redacted]	1.1	[redacted]
[redacted]	[redacted]	1.1	[redacted]
[redacted]	[redacted]	2.0	[redacted]
[redacted]	[redacted]	2.1	[redacted]
[redacted]	[redacted]	2.1	Genfit
[redacted]	[redacted]	2.1	Genfit
[redacted]	[redacted]	2.1	Genfit
[redacted]	[redacted]	2.1	Genfit
[redacted]	[redacted]	2.2	[redacted]
[redacted]	[redacted]	2.2	Genfit
[redacted]	[redacted]	2.2	Genfit
[redacted]	[redacted]	2.3	[redacted]
[redacted]	[redacted]	2.3	Genfit
[redacted]	[redacted]	2.3	Genfit
[redacted]	[redacted]	2.4	[redacted]
[redacted]	[redacted]	2.4	Genfit

[redacted]	[redacted]	2.4	Genfit
[redacted]	[redacted]	2.4	Genfit
[redacted]	[redacted]	2.4	Genfit
[redacted]	[redacted]	2.4	[redacted]
[redacted]	[redacted]	3.0	[redacted]
[redacted]	[redacted]	4.0	[redacted]
[redacted]	[redacted]	4.0	Genfit
[redacted]	[redacted]	4.0	Genfit
[redacted]	[redacted]	4.0	Genfit
[redacted]	[redacted]	4.0	Genfit
[redacted]	[redacted]	5.0	[redacted]
[redacted]	[redacted]	5.0	Genfit
[redacted]	[redacted]	5.0	Genfit
[redacted]	[redacted]	5.0	Genfit

VERSION HISTORY

Version Number	Version Date	Summary and rational of change(s)
2.0	26 Jan 2017	Update with protocol amendment 1 New section added with rules for dealing with missing or incomplete dates
3.0	17 Jan 2020	SAP split into two versions; for the surrogate endpoint analysis (SEA) and long-term (final) analysis Additional detail and clarifications provided for analysis set definitions (ITT F1 set, FSS, SS and SSF1) Additional section regarding estimand definitions Addition of new secondary endpoints Clarifications of time points, visit windows and handling of missing data Clarification and addition of sensitivity and supplementary analyses of time to first clinical event / death for final analysis Method of pooling centers clarified (sensitivity analysis including center as fixed effect) Update of the statistical method for the analysis of the binary secondary endpoints defined according to histological parameters. These analyses will be performed using the approach described in Ge et al. (2011) Addition of descriptive statistics on baseline characteristics, time to study withdrawal, treatment exposure and treatment emergent adverse events for patients with missing outcome for time to clinical event/death due to early withdrawal, and for patients with early study treatment discontinuation Additional compliance summary up to week 72 for SEA Addition of shift tables from baseline to week 72 and follow-up for fibrosis stage, ballooning, inflammation, steatosis, NAS disease activity score and SAF activity score Additional information provided regarding summaries of morphometry data Analyses of change from baseline in NAS disease activity score and Framingham risk scores aligned to have consistent assumptions of variable distribution at SEA and final Analysis of hsCRP and NT-ProBNP updated to be performed on log transformed values

		Addition of 2 subgroup analyses for primary endpoint Addition of exploratory analyses for the ITT F1 population Addition of exploratory analyses for the FITT population Addition of safety analyses for Safety F1 population Updated the adaptive design methodology Added raw counts of number of events for summaries of AE data Added clarification regarding which laboratory parameters will be analysed using repeated measures ANCOVA Added definition and analysis of eGFR New figures for laboratory parameters Updated section 8. "Changes in planned analyses" according to the protocol amendments Editorial changes
4.0	22 Apr 2020	Changes in the definition of the long-term primary endpoint Added definition and analysis of AESIs New figure for albuminuria Added clarifications regarding the strategy for eDISH plots Added clarifications regarding the categorization of subjects with serum creatinine increases Added analysis of eGFR decreases Removed the criteria for DILI adjudication that were not consistent with the protocol.
5.0	23 Nov 2020	Updated SAP to reflect change in final reporting specifications following sponsor's decision to terminate the trial prematurely. Details of the updates are contained in Section 7 Changes in planned analyses Clarifications and Editorial changes

TABLE OF CONTENTS

GLOSSARY OF ABBREVIATIONS

Abbreviation	Term
AE	Adverse Event
AESI	Adverse Event of Special Interest
AKI	Acute Kidney Injury
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass Index
BP	Blood Pressure
CEC	Clinical Events Committee
CI	Confidence Interval
COVID-19	Coronavirus Disease 2019
eCRF	Electronic Case Report Form
eGFR	Estimated Glomerular Filtration Rate
CRN	Clinical Research Network
DILI	Drug Induced Liver Injury
DSMB	Data Safety Monitoring Board
ECG	Electrocardiogram
FITT	Full Intent-to-treat
FSS	Full Safety Set
F1	Fibrosis Stage 1
F2	Fibrosis Stage 2
F3	Fibrosis Stage 3
GLP	Glucagon-Like Peptide
KDIGO	Kidney Disease Improving Global Outcomes
LTTP	Long Term Treatment Period
MedDRA	Medical Dictionary for Regulatory Activities
MELD	Model End Stage Liver Disease
NAFLD	Nonalcoholic Fatty Liver Disease
NAS	NAFLD Activity Score
NASH	Nonalcoholic Steatohepatitis
PT	Preferred Term
SAE	Serious Adverse Event
SAF	Steatosis-Activity-Fibrosis
SD	Standard Deviation
SE	Standard Error
SEA	Surrogate Endpoint Analysis
SF-36	36-Item Short-Form Health Survey
SOC	System Organ Class
SGLT	Sodium/Glucose Cotransporter

1 SOURCE DOCUMENTS

The Statistical Analysis Plan was written based on the following documentation:

Document	Date	Version
Protocol	15-Jan-2016	1.0
Protocol Amendment	06-Jan-2017 03-Apr-2017 03-Jan-2020 20-Apr-2020	2.0 3.0 4.0 5.0
Protocol specific for France	01-Jul-2016	1.1
Protocol Amendment specific for France	12-Jan-2017 10-Apr-2017 03-Jan-2020 23-Apr-2020	2.1 3.1 4.1 5.1
eCRF	12-May-2016 31-Aug-2016 05-Apr-2019	1.0 2.0 8.0
DSMB Charter	03-May-2016 10-Nov-2016 17-Feb-2020	1.0 2.0 3.0

This Final Analysis Statistical Analysis Plan details objectives, definitions and planned analysis for the final analysis of the study. For the same details regarding the surrogate endpoint analysis, see the separate Surrogate Endpoint Analysis Statistical Analysis Plan.

2 PROTOCOL DETAILS

Below follows the design and objectives defined in the protocol. Note that as a consequence of the sponsor's decision to terminate the trial prematurely, the planned final analyses differ from those set out in the protocol. Details of the updates are reflected in sections 3, 4, 5 and 6 of this SAP, and summarised in Section 7 Changes in planned analyses.

2.1 Study Objectives

2.1.1 Primary objective

The primary objective of the study is to evaluate the efficacy of elafibranor 120 mg versus placebo on resolution of Nonalcoholic Steatohepatitis (NASH) without worsening of fibrosis in patients with fibrosis stage 2 (F2) and fibrosis stage 3 (F3).

Primary objective – long-term endpoints (at final analysis)

To evaluate the efficacy of elafibranor 120 mg QD versus placebo on clinical outcomes described as a composite endpoint composed of death due to any cause, histological liver cirrhosis, and the full list of portal hypertension/cirrhosis related events:

- Liver transplantation
- Model end stage liver disease (MELD) score ≥15 for patients with baseline MELD score ≤ 12
- Onset of:

- variceal bleed requiring hospitalization,
- hepatic encephalopathy defined as West Haven/Conn score ≥ 2 and requiring hospitalization,
- spontaneous bacterial peritonitis,
- ascites requiring treatment,

2.1.2 Secondary objectives

Secondary objectives

- To assess histological changes after 72 weeks of treatment and at the end of the Long Term Treatment Period (LTTP) on the following endpoints:
  - percentage of patients with resolution of NASH without worsening of fibrosis
  - percentage of patients with improvement of fibrosis of at least 1 stage according to NASH clinical research network (CRN) scoring
  - percentage of patients with improvement of fibrosis of at least 1 stage according to NASH CRN scoring without worsening of NASH
  - percentage of patients with no worsening of Fibrosis and no worsening of NASH
  - percentage of patients with Resolution of NASH and improvement of Fibrosis
  - percentage of patients with at least a 1 point improvement in histological scores (NASH CRN scoring: NAS [sum of steatosis, hepatic ballooning and lobular inflammation], steatosis, hepatic ballooning, lobular inflammation), fibrosis (Nonalcoholic fatty liver disease [NAFLD] Ishak scoring system), or portal inflammation
  - percentage of patients with improvement of NAS of at least 2 points
  - percentage of patients with improvement of NAS of at least 2 points and with at least 1 point improvement in hepatic ballooning
  - percentage of patients with at least a 1 point improvement in disease activity score (sum of ballooning and lobular inflammation scores) according to NAS scoring and steatosis-activity-fibrosis (SAF) scoring
  - percentage of patients with at least a 1 point improvement in disease activity score (sum of ballooning and lobular inflammation scores) according to NAS scoring and with at least 1 point improvement in hepatic ballooning
  - percentage of patients with at least a 1 point improvement in disease activity score (sum of ballooning and lobular inflammation scores) according to steatosis-activity-fibrosis (SAF) scoring and with at least 1 point improvement in hepatic ballooning
  - percentage of patients with at least 2 points improvement in disease activity score according to NAS scoring and SAF scoring
  - changes in NAS, fibrosis (using NASH CRN and NAFLD Ishak scoring system), steatosis, hepatic ballooning, lobular inflammation, portal inflammation, SAF activity score
  - changes in area of fibrosis (normalized Collagen Proportional area in %) by morphometry
- To assess the following endpoints at Week 72, and at the end of the LTTP:
  - changes in liver enzymes and liver markers
  - changes in noninvasive markers of fibrosis and steatosis
  - changes in lipid parameters
  - variation in body weight
  - changes in insulin resistance and glucose homeostasis markers

- changes in inflammatory markers
- changes in cardiovascular risk profile as assessed by Framingham scores
- changes in liver stiffness by Fibroscan measurement
- changes in quality of life (36-Item Short-Form Health Survey (SF-36) questionnaire)

- To assess the onset of:
  - histological liver cirrhosis
  - death of any cause
  - any portal hypertension or cirrhosis related events
  - cardiovascular events
  - liver-related death events.

Safety secondary objectives

- To assess the tolerability and safety of once a day administration of oral doses of elafibranor 120 mg:
  - adverse events
  - clinical laboratory tests (hematology, chemistry and urinalysis)
  - vital signs
  - electrocardiogram
  - physical examination
  - renal, cardiac, liver, and metabolic parameters
  - other safety markers.

2.1.3 Exploratory objectives

Exploratory objectives

- To constitute a biobank for discovery and validation of biomarkers in NASH.

Exploratory objectives for F1 group

An exploratory objective of the study is to evaluate the efficacy of elafibranor 120 mg versus placebo in a subset of patients with fibrosis stage 1 (F1). The same efficacy and safety endpoints as for the analysis on the F2/F3 cohort will be assessed.

2.2 Overall Study Design

This is a multicenter, randomized, double-blind, placebo-controlled phase III study in patients with NASH.

There are two treatment arms:

- Elafibranor 120 mg
- Placebo

Treatment will be given by oral administration once per day.

Random allocation will be made to the two treatment groups in a 2:1 ratio (elafibranor:placebo) and stratified by the following factors:

- Type 2 diabetes (yes/no)
- Gender (male/female)
- Fibrosis stage (F1/F2/F3)

There will be a 12 week screening period. The first treatment period will last 72 weeks with visits every 12 weeks, and will be followed by a LTTP with visits every 24 weeks. The duration of the LTTP will be driven by the number of events and is anticipated to last for approximately 4 years.

The primary comparison of elafibranor versus placebo will be based on a subset of F2 and F3 patients. An additional exploratory comparison will include the subset of F1 patients.

2.3 Sample Size and Power (Final Analysis)

2.3.1 Time to clinical event/death

The following assumptions were made for the sample size calculation for time to clinical event/death:

- 24 month enrollment (with an 18-month ramp up to as many as 200 patients per month)
- 72 month maximum follow-up
- α=0.04 two sided
- Randomization ratio of 2:1 (elafibranor to placebo)
- 7% annual event rate in control group (placebo)
- Hazard ratio of 0.75 in favor of active group (elafibranor)
- 4% annual drop-out rate over 72 months

Based on these assumptions, 456 events are required to provide 80% power to show that elafibranor is superior to placebo with respect to time to clinical event/death. In order to obtain 456 events, at least 2022 patients will be required in the ITT population.

Sample size calculations were performed in EAST version 6.3.

2.4 Estimand strategy

In line with ICH E9(R1) addendum; five attributes (treatment condition, population, endpoint, intercurrent events and population level summary) have been specified to translate the primary efficacy objective into treatment effect that are to be estimated (estimands).

The intercurrent events as the non-adherence to study treatment, study treatment discontinuation due to adverse event, refusal of the patients to perform the biopsy, change in diet/exercise, change in dose or initiation of glucagon-like peptide-1 (GLP-1) agonists, sodium/glucose cotransporter-2 (SGLT-2) inhibitors, statins, new approved or investigational NASH therapies can be considered as intercurrent events that could alter the interpretation of the long term clinical outcome.

The primary estimand strategy will be based on the treatment policy strategy:

- *A. Treatment condition:* Elafibranor 120 mg administered daily in comparison with matching placebo 120 mg administered daily.
- *B. Population:* Randomized patients with fibrosis stage 2 (F2) or fibrosis stage 3 (F3), and NAS ≥4
- *C. Endpoint:* Time to occurrence of long-term composite endpoint,
- *D. Intercurrent events:* Regardless of the intercurrent events,
- *E. Population-level summary:* Between treatment group Hazard Ratio.

In addition, to take into account the treatment switching to new approved or investigational NASH therapies, the hypothetical strategy will be also investigated as supplementary estimands and sensitivity analysis imputing as if:

1) The patients would follow their initial randomized treatment
2) The patients from both arms would continue on placebo

3 EFFICACY AND SAFETY VARIABLES

3.1 Efficacy Variables

The definition of the efficacy and safety variables has been amended as a consequence of the sponsor's decision to terminate the trial prematurely. A summary of changes is included in Section 7.

Time to clinical event/death

The composite endpoint of time to clinical event/death is composed of death due to any cause, histological liver cirrhosis and the full list of portal hypertension/cirrhosis related events as follows:

- Liver transplantation
- MELD score ≥ 15 for patients with baseline MELD score ≤ 12
- Onset of:
  - variceal bleed requiring hospitalization
  - hepatic encephalopathy defined as West Haven/Conn score ≥ 2 and requiring hospitalization
  - spontaneous bacterial peritonitis
  - ascites requiring treatment

Time to onset of cardiovascular events

The composite endpoint of time to onset of cardiovascular events is composed of the following cardiovascular events:

- Non-fatal myocardial infarction/unstable angina
- Non-fatal stroke
- Unstable Angina
- Hospitalization for Heart Failure

- Coronary Revascularization (bypass or PCI)

3.2 Safety Variables

The tolerability and safety will be assessed using the following endpoints:

- Adverse events
- Clinical laboratory tests (hematology, chemistry and urinalysis)
- Vital signs
- Electrocardiogram
- Physical examination
- Renal, cardiac, liver, and metabolic parameters
- Other safety markers
- Drug induced liver injury (DILI)

4 ANALYSIS POPULATIONS

The definition of the analysis populations has been amended as a consequence of the sponsor's decision to terminate the trial prematurely. A summary of changes is included in Section 7.

4.1 Enrolled Set

All patients who signed informed consent.

4.2 Full Safety Set

All patients who have taken at least one dose of study treatment will be included in the Full Safety Set. Patients will be analyzed according to their actual treatment received.

4.3 Full Intent-to-treat Set

All randomized patients. Patients will be analyzed according to their randomized treatment.

5 DATA HANDLING

5.1 Time points and Visit Windows

The following rules will be used for the by-visit analyses of the weight, clinical laboratory tests, vital signs, electrocardiogram (ECG) and other safety markers.

For all populations, assessments will be assigned to visits for non-categorical summaries as follows:

- Only assessments recorded with a nominal visit number or recorded as an unscheduled visit will be considered for assignment to visits (i.e. results recorded on the end of treatment and end of study CRF forms will not be used).

- For baseline assessments, the latest assessment with non-missing result on or before study day 1 will be used. Note that this may include visits also assigned to screening for by-visit summaries. In the event that multiple assessments occur on the same day, a nominal visit 1 will be used in preference to a nominal screening visit.
- For post-baseline assessments, only those up to and including 30 days post last dose of study treatment will be considered for assignment to visits.
- Assessments with missing data and assessments marked 'Not Done' will be as providing a missing response and are not permitted to be assigned to a visit window.
- If the initial nominal visit, (defined hereafter as the earliest assessment for a given visit to be recorded as the nominal visit number via the CRF), has a non-missing response, then that will be chosen for analysis.
- If the initial nominal visit has a missing response, then the earliest of all subsequent assessments with a non-missing response which have either:
    - been recorded as the same nominal visit, or
    - been recorded as an unscheduled visit which is assigned to the given visit as per Table 1

  will be chosen for the analysis.
- If a given visit has no initial nominal visit record, then the earliest of all unscheduled visits which are assigned to the given visit as per Table 1 will be chosen for the analysis.
- If in the event that none of the above rules can choose between multiple assessments for a given visit, the following additional rules in the order below will be used:
    - For parameters with normal range limits, if one of the options is outside of the normal range, and the other(s) is/are not, the assessment outside of the normal range will be chosen.
    - For parameters with normal range limits, if multiple options are outside of the normal range, and all in the same direction (i.e. all below the normal range, or all above the normal range), the more extreme value (i.e. the lowest in the case of below the normal range, or the highest in the case of above the normal range) will be chosen.
    - If the above rules are unable to resolve an assignment to the visit, then the assessment with the lower record ID will nominally be chosen.

For all populations, assessments will be assigned to visits for safety categorical summaries (e.g. clinical laboratory data) as follows:

- Only assessments recorded with a nominal visit number or recorded as an unscheduled visit will be considered for assignment to visits.
- Assessments with missing data and assessments marked 'Not Done' will be as providing a missing response and are not permitted to be assigned to a visit window.
- The worse value (e.g. out of reference range preferred to within range; abnormal preferred to normal) will be used in each window

End of the 72 week treatment period visit will be defined as the last post-baseline visit during the 72 week double-blind treatment period.

End of the long term treatment period will be defined as the last post-baseline visit during the 72 week double-blind and long term treatment period.

Day 1 will be defined as the day of first dose of double-blind study treatment. For patients randomized not treated, Day 1 will be defined as the day of randomization.
Relative days after Day 1 are calculated as (assessment date – Day 1 date) + 1. Relative days prior to Day 1 are calculated as (assessment date – Day 1 date). The day prior to Day 1 is Day -1.

Table 1 Definition of visit windows

Visit	Target Day of Visit[a]	Acceptable visit window
1 Screening	Day –X[b]	Up to Day -1
2 Baseline	Day 1	*See baseline definition above
3 Week 12	Day 84	Days 2 to 125
4 Week 24	Day 168	Days 126 to 209
5 Week 36	Day 252	Days 210 to 293
6 Week 48	Day 336	Days 294 to 377
7 Week 60	Day 420	Days 378 to 461
8 Week 72	Day 504	Days 462 to 587
9 Week 96	Day 672	Days 588 to 755
10 Week 120	Day 840	Days 756 to 923
11 Week 144	Day 1008	Days 924 to 1091
12 Week 168	Day 1176	Days 1092 to 1259
13 Week 192	Day 1344	Days 1260 to 1427
14 Week 216	Day 1512	Days 1428 to 1595
15 Week 240	Day 1680	Days 1596 to 1763
16 Week 264	Day 1848	Days 1764 to 1931
17 Week 288	Day 2016	Days 1932 to 2099
18 Week 312	Day 2184	Days 2100 to 2267
End of the 72 week treatment period	N/A	Last post-baseline visit during the 72 week double-blind treatment period
End of the long term treatment period	N/A	Last post-baseline visit during the 72 week double-blind and long term treatment period

[a] relative to the date of first dose of double-blind study treatment (Day 1)
[b] screening visit has different target days depending upon the assessment performed

All data will be listed. However, only data slotted into the protocol defined visits will be reported in table summaries. For example, ECG data will only be summarized at Baseline, Week 36, Week 72, Week 120 etc.

All other analysis will use the nominal study visit as defined in the Study Schedule and eCRF.

5.2 Handling of Dropouts or Missing Data

5.2.1 Handling of Missing Data

Missing outcome data for time to event endpoints (e.g. endpoint of time to clinical event/death) will be censored at the last known date for the analysis.

5.2.2 Handling of Missing or Incomplete Dates

Incomplete dates (partial or missing dates) will be presented in the data listings as provided on the electronic Case Report Form (eCRF). However, for use in calculations (e.g. to calculate the duration of an AE or medication use) dates will be estimated as follows:

For partial start dates:

If the year is unknown, then:

- The date will not be imputed and will be assigned a missing value.

If the month is unknown, then:

- If the year matches the year of the first dose date, then impute the month and day of the first dose date.
- Otherwise assign "January".

If the day is unknown, then:

- If the month and year match the month and year of the first dose date, then impute the day of the first dose date.
- Otherwise assign "01".

For partial end dates:

- If the year is unknown, then the date will not be imputed and will be assigned a missing value.
- If the month is unknown, then assign "December".
- If the day is unknown, then assign the last day of the month.

If the above rules for end-dates result in illogical date with regards to the dates the subject was in the study, then the end date will be replaced with the subject's date of completion/discontinuation.

6 STATISTICAL METHODS

The definition of the statistical methods has been amended as a consequence of the sponsor's decision to terminate the trial prematurely. A summary of changes is included in Section 7.

6.1 Final Analysis

The final analysis will take place at the time of final database lock.

At this time, the database will be cleaned and locked for analysis.

All safety data collected will be summarized in safety outputs.

6.2 General Principles

All data processing, summarization and analyses will be performed using the Hosted SAS Environment / Version 9.1.3 (or later) of the SAS® statistical software package.

Unless specified otherwise, data will be displayed using the following treatment group labels, in the order presented:

- Elafibranor
- Placebo

All data collected will be presented in listings by treatment group, country, patient, assessment and visit (where applicable), unless otherwise specified.

Data will be presented in summary tables by treatment group, assessment and visit (where applicable).

The category "Missing" will be presented if the number missing is greater than zero for at least one treatment group.

Descriptive summary statistics for continuous variables will include the number of patients (n), mean, standard deviation (SD), median, minimum and maximum.

In order to calculate summary statistics for log-transformed continuous variables, results will be log-transformed at the patient level and summary statistics will be calculated after back-transformation to provide the geometric mean and associated standard error (SE). Descriptive summary statistics for log-transformed continuous variables will include the n, geometric mean, SE, median, minimum and maximum.

Descriptive summary statistics for categorical variables will include frequency counts and percentages [n (%)]. Unless stated otherwise in the table shells, the denominator for percentage calculations will be the number of patients in the analysis population.

Change from baseline will be calculated where both baseline and corresponding post-baseline assessments are present.

Dates will be displayed as DDMMMYYYY.

6.3 Patient Disposition and Data Sets Analysed

Patient disposition will be listed and summarized by treatment group and overall.

The number and percentage of patients in the following categories will be summarized for patients in the enrolled population and in the subset of patients from the enrolled population with confirmed or suspected coronavirus disease 2019 (COVID-19) (see section 6.8.2 for identification of such patients):

- screened
- randomized
- non-randomized (screen failures)

- randomized and not treated
- treated

The number and percentage of patients in each of the analysis populations (Enrolled, full intent-to-treat [FITT] and full safety set [FSS]) will be summarized by treatment group and overall for patients in the enrolled population.

The number and percentage of patients who complete the study and who discontinue the study early, including a breakdown of the primary reasons for discontinuation, will be presented for the FSS population. In addition, the number and percentage of patients who complete the treatment and who discontinue the treatment early, including a breakdown of the primary reasons for discontinuation of treatment early will be presented.

The primary reasons for both treatment and study discontinuation will be supplemented with a breakdown of patients discontinuing for reasons related to COVID-19.

- patients will be deemed to have discontinued treatment due to COVID-19 if they have a record of treatment discontinuation due to an AE, which itself has the AE identification number linked to an AE which has been identified as COVID-19 related as per the rules in section 6.8.2.
- Patients will be deemed to have discontinued study due to COVID-19 if they have a reference to COVID-19 in free text field entered within the study discontinuation form.

In the subset of patients from the FSS with confirmed or suspected COVID-19, the study completion status of patients with the primary reasons for discontinuation will be displayed.

A summary of patient enrollment by country and site will also be provided by treatment group and overall for the FSS population.

A summary of the reasons for screen failure will be produced for patients in the enrolled population. No other information for screen failures will be presented.

A summary of patients who attended each visit will be produced for the FSS population.

6.4 Protocol Deviations

All important protocol deviations will be listed for the FITT population.

All important protocol deviations will be summarized by treatment group for the FITT and FSS populations. An additional summary of all important deviations related to COVID-19 (identified as those entered as related to COVID-19 in the dedicated associated free text field) will be presented in the same manner for the FSS population.

The important protocol deviations will be identified before data are unblinded.

6.5 Demographics and Other Baseline Characteristics

Demographic and baseline characteristics will be listed and summarized by treatment group and overall for patients in the FITT and FSS populations.

These characteristics will also be described in:

- the subset of patients from the FSS set with confirmed or suspected COVID (identified as per the rules in section 6.8.2).

Standard descriptive statistics will be presented for the continuous variables of:

- age (years);
- weight (kg);
- height (cm);
- body mass index (kg/m^2) [calculated as ($weight/height^2$) where weight is in kg and height is in m];
- waist circumference (cm);
- NAS (calculated as the sum of steatosis, hepatic ballooning and lobular inflammation);
- NAFLD fibrosis score;
- Model end stage liver disease score (MELD) score;

The total counts and percentages of patients will be presented for the categorical variables of:

- age group (years) (grouped as <60, and ≥60);
- gender (male, female);
- race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Pacific Islander, White);
- ethnicity (Hispanic or Latino, Not Hispanic or Latino);
- type 2 diabetes (yes, no);
- fibrosis stage (F1, F2, F3);
- child bearing potential (yes, no, not applicable);
- NAS severity [grouped as moderate (4-5), and severe (≥6)];
- MELD score (<15, ≥15);

No formal tests of statistical significance will be performed on the demographic and baseline data.

Other baseline measurements, such as vital signs, will be summarized by treatment group with the post-baseline measurements.

6.5.1 Medical History

Medical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) [Version 19.0 (or a later version if updated during the study)]. All medical history will be listed, and the number and percentage of patients with any medical history will be summarized for patients in the FSS population by system organ class (SOC) and preferred term (PT) for each treatment group and overall.

6.5.2 Prior and Concomitant Medications

Medications received prior to or concomitantly with study treatment will be coded by [redacted] using the WHO Drug Dictionary [Version B2-BDE March 2016 (or a later version if updated during the study)], Anatomical Therapeutic Chemical (ATC) Classification codes.

Prior medications and concomitant medications are defined as follows:

Prior medications are those taken with a stop date prior to the first dose date of study treatment.

Concomitant medications are those with a start date on or after the first dose date of study treatment, or those with a start date before the first dose date of study treatment and a stop date on or after the first dose date of study treatment.

If a medication cannot be classified as "prior" or "concomitant" after applying imputation rules for missing/incomplete dates, it will be classified as concomitant.

Prior medications and concomitant medications will be listed together and summarized separately for patients in the FSS population.

The number and percentage of patients using each medication will be displayed together with the number and percentage of patients using at least one medication within each therapeutic class (ATC-Level 2), chemical subgroup (ATC-Level 4), and generic term.

6.5.3 Substance Use History

Substance use history will be listed and summarized by treatment group and overall for patients in the FSS population.

The total counts and percentages of patients will be presented for the categorical variables of:

- Does the patient smoke or have a history of smoking? (Yes, No)
- Smoking status (Current, Former)
- If current or former smoker, type of substance (Cigarettes, Pipes, Cigars, Other)
- Does the patient consume alcohol? (Yes, No)
- If consume alcohol, frequency of alcohol consumption (< 1-7 drink units per week, 8 drink units per week, 9-14 drink units per week, 15 or more drink units per week)
- Does the patient consume caffeinated coffee? (Yes, No)
- Frequency of coffee consumption (< 1-7 cups per week, 8 cups per week, 9-14 cups per week, 15 or more cups per week)

In addition, the number of cigarettes / pipes / cigars smoked per day will be summarised for current smokers.

6.5.4 Safety, Diet and Lifestyle Factors

Safety Diet and Lifestyle Factors will be listed and summarized by visit, treatment group and overall for patients in the FSS population.

The total counts and percentages of patients will be presented for the categorical variables of:

- Significant changes in diet since last visit? (Yes, No)
- Significant change in physical activity since last visit? (Yes, increased, Yes, decreased, No)
- Does the patient consume alcohol? (Yes, No)
- Type of alcohol consumption, (Beer, Wine, Distilled spirits or liquor, Other)
- Frequency of alcohol consumption (< 1-7 drink units per week, 8 drink units per week, 9-14 drink units per week, 15 or more drink units per week, Other)
- Significant change in dietary supplement intake? (Yes, No)
- Does the patient consume caffeinated coffee? (Yes, No)
- Frequency of coffee consumption (< 1-7 cups per week, 8 cups per week, 9-14 cups per week, 15 or more cups per week)
- Current smoking status (Smoker, Non-smoker)
- Occurrence of diabetes? (Yes, No)
- Currently on medication for high blood pressure (Yes, No)
- Patient still compliant with study drug intake (Yes, No, Not applicable)

6.6 Measurements of Treatment Compliance

Percentage compliance will be calculated as:

100 * actual tablets taken/expected tablets taken

Where actual tablets taken is defined as the sum of the tablets taken and tablets expected to be taken is defined as (date of last dose – date of first dose) + 1.

Percentage compliance will be summarized descriptively by treatment group for patients in the FITT and FSS populations.

The number and percentage of compliant patients will be presented for the FITT and FSS populations, where compliant is defined as percentage compliance between 80.0% and 120.0% inclusive. The following percentage compliance categories will be presented:

- <50.0%
- ≥50.0% to <80.0%
- ≥80.0% to ≤120.0%
- >120.0%

6.7 Efficacy

6.7.1 Efficacy Analysis

The analysis of the efficacy endpoints will be performed on the FITT population.

6.7.1.1 Clinical events / death

The composite endpoint of time to clinical event/death is composed of the following:

- Death to any cause
- Histological liver cirrhosis
- Portal hypertension/cirrhosis related events:
    - Liver transplantation
    - MELD score $\geq$ 15 for patients with baseline MELD score $\leq$ 12
    - Onset of:
        i. variceal bleed requiring hospitalization
        ii. hepatic encephalopathy defined as West Haven/Conn score $\geq$ 2 and requiring hospitalization
        iii. spontaneous bacterial peritonitis
        iv. ascites requiring treatment.

All events must be adjudicated by the Clinical Events Committee (CEC) to meet the protocol-specified criteria in order to contribute towards the composite endpoint. The CEC assessment and adjudication will occur in a blinded, consistent, and unbiased manner throughout and the CEC's processes and responsibilities are defined within the CEC charter.

The number and percentage of patients with each type of efficacy related adjudicated event will be summarized by treatment group. Of these events, the number and percentage of patients with liver related death events will also be presented within this summary.

The time to first clinical event/death (months) will be calculated for each patient as follows:

(date of first clinical event/death – date of first dose of study medication) + 1)/30.4375

Patients that do not have a clinical event/death will be censored at the last known date.

The time to first clinical event/death will be analyzed using a Cox proportional hazard's model with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female), and fibrosis stage (F2, F3).

The analysis will be performed using 'PROC PHREG'. An example of the SAS code for the Cox proportional hazard's model is below:

```
proc phreg data = <dataset>;
  class trtpn diab gender fibrosis;
  model tevent*cnsr(1) = trtpn diab gender fibrosis;
  hazardratio trtpn;
run;
```

where

tevent = time to first event (months)

cnsr = event(0), censored (1)
trtpn = treatment (numeric value for planned treatment)
diab = type 2 diabetes category (yes, no)
gender = gender category (male, female)
fibrosis = fibrosis stage (F2, F3)

The statistical model will be used to calculate the adjusted hazard ratio (elafibranor/placebo) and associated 95% CI to be presented.

The time to first clinical event/death will be presented graphically using a Kaplan-Meier curve.

A summary of estimated survival probabilities with 95% confidence intervals (CIs) using the Kaplan-Meier method will be presented at 6 month intervals until 6 years. The number of patients under risk, number of events and censored will also be included. Furthermore, the estimated median time to first clinical event/death and 95% CI will be presented for each treatment group.

6.7.1.2 Cardiovascular events

Occurrence of cardiovascular events will be assessed and adjudicated by an independent CEC.

The following cardiovascular events will be assessed:

- Non-fatal myocardial infarction/unstable angina
- Non-fatal stroke
- Unstable Angina
- Hospitalization for Heart Failure
- Coronary Revascularization (bypass or PCI)

The time to first cardiovascular event (months) will be calculated for each patient as follows:

(date of first cardiovascular event – date of first dose of study medication) + 1)/30.4375

Patients that do not have a cardiovascular event will be censored at the last known date.

The time to first cardiovascular event will be analysed using a Cox proportional hazard's model with fixed terms for treatment, type 2 diabetes (yes, no), gender (male, female) and fibrosis stage (F2, F3). The statistical model will be used to calculate the hazard ratio (elafibranor/placebo) and 95% CI. The hazard ratio and 95% CI will be presented.
The time to first cardiovascular event will be presented graphically using a Kaplan-Meier curve.

A summary of estimated survival probabilities with 95% CIs using the Kaplan-Meier method will be presented at 6 monthly intervals until 6 years. The number of patients under risk, number of events and censored will also be included. Furthermore, the estimated median time to first cardiovascular event and 95% CI will be presented for each treatment group.

In addition, the number and percentage of patients with each type of adjudicated cardiovascular event will be summarized by treatment group.

6.8 Safety

Unless otherwise stated, all safety summaries will be presented for patients in the FSS population.

6.8.1 Extent of Exposure

Duration of exposure will be defined in weeks as:

([last dose date – first dose date] + 1– off-drug days) / 7

If date of first dose date is missing, then date of randomization visit will be used. If last dose date is missing, then date of last known dose will be used.

Duration of exposure will be listed and summarized using descriptive statistics for each treatment group, and overall.

The number and percentage of patients with duration of exposure in the following categories will be summarized:

- ≥0 and <12 weeks
- ≥12 and <24 weeks
- ≥24 and <36 weeks
- ≥36 and <48 weeks
- ≥48 and <60 weeks
- ≥60 and <72 weeks
- ≥72 and <96 weeks
- ≥96 and <120 weeks
- ≥120 and <144 weeks
- ≥144 and <168 weeks
- ≥168 and <192 weeks
- ≥192 and <216 weeks
- ≥216 and <240 weeks
- ≥240 and <264 weeks
- ≥264 and <288 weeks
- ≥288 and <312 weeks
- ≥312 weeks

6.8.2 Adverse Events

All adverse events (AEs) recorded on the eCRF will be coded using the MedDRA dictionary [Version 23.0 (or a later version if updated during the study)] and classified as either non treatment-emergent AEs (pre-treatment AEs, post-treatment AEs) or treatment emergent AEs (TEAEs) as follows:

- Pre-treatment AEs are events that start prior to the date of first dose of study treatment.
- Post-treatment AEs are events that start more than 30 days after stop of study treatment.
- TEAEs are events with start date on or after the date of first dose of study treatment and up to 30 days after date of last dose of study treatment or events with start date prior to the date of first dose of study treatment whose severity worsens on or after the date of first dose of study treatment.

All AE data will be listed by treatment group. Treatment-emergence status will be flagged in the listing. In addition, corresponding listings of serious AEs (SAEs), AEs leading to discontinuation of study treatment and AEs resulting in death will be produced.

The number and percentage of patients reporting each pre-treatment AE will be summarized for each treatment group and overall, by SOC (sorted alphabetically) and PT (sorted by descending overall total).

Summary tables of TEAEs by treatment group and overall will be produced.

The severity of all AEs is recorded as mild, moderate, or severe. If severity is missing for a TEAE, it will be considered severe only in the overall category in the summary tables.

The relationship between an AE and treatment is assessed as 'not related, unlikely related, possibly related, related, and not assessable'. A treatment-related AE is an AE considered by the investigator as possibly related, related to treatment, and not assessable, or with unknown/missing relationship to treatment.

Standardised MedDRA queries (SMQs) will be used to compare each AE against a pre-defined list of COVID-19 related terms to classify each AE as COVID-19 related or non-COVID-19 related. The full list of the SMQs will be approved and signed off in a separate document by Genfit prior to database lock. Patients will also be flagged as having either confirmed or suspected COVID-19 if they have at least one such event.

An overview table will summarize the number and percentage of patients with at least one of the following TEAEs, where patients with more than one TEAE in a particular category are counted only once in that category:

- any AE;
- any AE by severity (mild, moderate, severe);
- treatment-related AE;
- AE leading to treatment discontinuation;
- treatment related AE leading to discontinuation;
- SAE;
- treatment-related SAE;
- SAE leading to death;
- treatment-related SAE leading to death;
- SAE leading to treatment discontinuation;
- any COVID-19 related AE
- any COVID-19 related AE leading to treatment discontinuation
- any serious COVID-19 related AE
- any serious COVID-19 related AE leading to treatment discontinuation

- any serious COVID-19 related AE leading to death.

The overview table will also summarize the number of the following TEAEs:

- all AEs;
- AEs by severity (mild, moderate, severe);
- treatment-related AEs;
- treatment-related AEs by severity (mild, moderate, severe);
- SAEs;
- SAEs by severity (mild, moderate, severe);
- treatment-related SAEs
- COVID-19 related AEs
- COVID-19 related AEs leading to treatment discontinuation
- serious COVID-19 related AEs
- serious COVID-19 related AEs leading to treatment discontinuation
- serious COVID-19 related AEs leading to death.

The number and percentage of patients reporting each TEAE and the total count of TEAEs will be summarized by SOC and PT. Tables will be sorted alphabetically by SOC. PTs will be sorted by descending frequency overall total. The following summaries will be produced:

- AEs by SOC and PT;
- Most common AEs (≥ 5% in any treatment group) by SOC and PT;
- AEs related to treatment, by SOC and PT;
- AEs by relationship to treatment, by SOC and PT;
- AEs by maximum severity, by SOC and PT;
- AEs related to treatment by maximum severity, by SOC and PT;
- AEs causing discontinuation from treatment, by SOC and PT;
- AEs related to study treatment causing discontinuation from treatment, by SOC and PT;
- AEs causing treatment delays and interruptions, by SOC and PT;
- SAEs, by SOC and PT;
- SAEs related to treatment, by SOC and PT;
- AEs related to COVID-19, by SOC and PT;
- AEs leading to death, by SOC and PT.

In the above summaries, patients with more than one AE within a particular SOC are counted only once for that SOC. Similarly, patients with more than one AE within a particular PT are counted only once for that PT. For summaries by maximum severity, patients with multiple AEs within a particular SOC or PT will be counted under the category of their most severe AE within that SOC or PT. AEs with missing intensity/severity will be included (as severe) in the overall count of patients with AEs.

AEs of special interest (AESI) are TEAEs which are defined according to categories and sub-categories as follows:

- CPK elevations of severe intensity or leading to permanent study drug discontinuation
- Muscle injury symptoms of severe intensity corresponding to:
  - Muscle pain or Myalgia

  - Muscle spasms or Tremor
  - Muscle weakness
- Transaminases elevations from baseline of severe intensity or leading to permanent study drug discontinuation
- Liver injury events of severe intensity corresponding to:
  - Hepatic impairment
  - Hepatic failure
- Gastrointestinal symptoms of severe intensity corresponding to:
  - Abdominal pain
  - Constipation
  - Diarrhea
  - Nausea
  - Vomiting
  - Acute cholecystis
  - Acute pancreatitis
- Fatigue and Asthenia of severe intensity
- Serum creatinine elevations of severe intensity or leading to permanent study drug discontinuation
- Renal injury events of moderate or severe intensity corresponding to:
  - Renal failure
  - Renal impairment
  - Renal colic
- Treatment emergent Pregnancy and outcomes of Pregnancy

All of the above categories, with the exception of the pregnancy AESIs, will be identified via a list of MedDRA queries of PT codes. Pregnancy AESIs will be identified using positive pregnancy assessment and reported in the pregnancy report form.

For each of the AESI categories and subcategories defined above, besides treatment emergent Pregnancy, the number and percentage of patients with at least one event and the corresponding number of events will be summarized for any:

- AESIs;
- Treatment related AESIs;
- Serious AESIs;
- AESIs leading to death;
- AESIs leading to treatment discontinuation.

In addition, the total number and percentage of patients with at least one AESI (with the exception of treatment emergent Pregnancy), and the corresponding number of events will be summarized for the five categories above.

For Pregnancy, the number and percentage of patients with at least one event and the corresponding number of events will be summarized. The outcomes of Pregnancy will also be summarized by SOC and PT.

In addition, the following summaries will be produced for the post-treatment AEs:

- Overall summary of post-treatment AEs;
- Overall summary of number of post-treatment AEs;
- AEs by SOC and PT;
- SAEs by SOC and PT.

No statistical comparisons of AEs between treatment groups will be performed.

6.8.3 Laboratory Evaluations

Clinical laboratory data to be summarized includes hematology, blood chemistry, and urinalysis parameters received from the central laboratory.

The table below presents the clinical laboratory tests to be performed as specified in the protocol.

Table 2 Clinical Laboratory Tests

Clinical Laboratory Test	Parameter
Hematology	Differential Count Hemoglobin Hematocrit Platelet Count PT (INR) Red Blood Cell Count Reticulocytes White Blood Cell Count
Serum Chemistry	Alanine aminotransferase (ALT) Albumin Alkaline phosphatase Apolipopprotein (ApoAI and ApoB) Aspartate aminotransferase (AST) Blood Urea Nitrogen Calcium Chloride Creatinine Creatine phosphokinase Direct Bilirubin (Conjugated Bilirubin in the protocol) Estimated glomerular filtration rate [MDRD, CKD-EPI (Creatinine), CKD-EPI (Cystatin C)][a] Fasting plasma glucose Gamma-Glutamyltransferase (GGT) Glucose Insulin Hemoglobin A1c High-sensitivity C-reactive protein

	High density lipoprotein-C Homeostasis model assessment of insulin resistance Low density lipoprotein-C Non High density lipoprotein-C Triglycerides Total cholesterol Total Bilirubin Very low density lipoprotein-C
Urinalysis	Albumin Albumin-creatinine ratio Creatinine Interleukin-18 Kidney injury molecule-1 Microscopic analysis a1 microgobulin Neutrophil gelatinase-associated lipocalin (NGAL) N-Acetyl-β-glucosaminidase (beta-NAG)
Urinalysis (dipstick)	Bilirubin Blood Glucose Ketones Leukocytes Nitrite pH Protein Specific gravity Urobilinogen
[a] Estimate glomerular filtration rate (mL/min/1.73m^2) is derived as: 133 x [minimum of Standardized serum cystatin C [mg/l] / 0.8, 1]$^{-0.499}$ x [maximum of Standardized serum cystatin C [mg/l] / 0.8, 1]$^{-1.328}$ x 0.996$^{Age\ (years)}$ x 0.932 [if female]	

All laboratory data will be reported in International System of Units (SI) units.

All laboratory data will be listed. Out-of-reference-range values will be flagged as high (H) or low (L) in the listings.

For analysis purposes, values preceded by a "<" or a ">" sign (i.e. those below or above the limits of quantification) will be considered equal to the lower or upper limit of quantification, respectively.

Only clinical laboratory tests included in the table will be included in the summary tables.

Clinical laboratory values recorded at each time-point (baseline, week 12, week 24, week 48, week 72, week 96, at every subsequent 24 weeks, at the end of the 72 week treatment period and at the end of the LTTP) and change from baseline to each post baseline time-point will be summarized by treatment group and overall as continuous data.

Line plots of means with whiskers extending to the mean +/ SE over time using the same visits as for table summaries will be presented for hemoglobin, hematocrit, RBC count, ALT, AST, GGT, alkaline phosphatase, creatinine, non high density lipoprotein-C, high density lipoprotein-C, low density lipoprotein-C, very low density lipoprotein-C, triglycerides, total cholesterol, a1 microgobulin, beta-NAG, NGAL, urinary creatinine and urinary albumin-creatinine ratio.

For hematology and serum chemistry, shift tables presenting movement in and out of reference range from baseline to end of the 72 week treatment period and at the end of the LTTP will be provided for each treatment group. Corresponding shift tables normal vs. abnormal will be produced for quantitative urinalysis parameters.

eDISH plots will present the concomitant values of aminotransferase (ALT and AST) and total bilirubin at the visit corresponding to the peak of aminotransferase (where scheduled visits are assigned to their nominal entered visit value, and unscheduled visits are assigned to visits according to Table 1). Distinct plots will be produced separately with regards to the peaks of ALT and AST. For each subject and for both ALT and AST, the strategy is as follows:

1) Among all the visits, identify the maximum value of aminotransferase
2) If the peak value occurs at just one visit, keep the highest value of total bilirubin at the visit corresponding to the peak of aminotransferase.
3) If the peak value occurs at more than one visit, keep the highest value of total bilirubin at any of the visits corresponding to the peak of aminotransferase
4) If the peak value occurs at a visit / visits with no corresponding value/values of total bilirubin, keep the value of bilirubin closest to the corresponding visit / any of the corresponding visits. In the event of equidistant time from the peak value visit / visits, keep the highest value of total bilirubin at any of the visits equidistant from the corresponding visit/visits of peak aminotransferase
5) Classify subjects into two subgroups: baseline value of ALT < ULN or baseline value of ALT ≥ ULN.
   a. If baseline value of ALT < ULN, express total bilirubin in xULN on the y-axis and aminotransferase in xULN on the x-axis. Add horizontal and vertical lines corresponding to total bilirubin = 2 ULN and aminotransferase = 3 ULN respectively.
   b. If baseline value of ALT ≥ ULN, present two plots:
      i. Express total bilirubin in xBaseline on the y-axis and aminotransferase in xBaseline on the x-axis. Add horizontal and vertical lines corresponding to total bilirubin = 2xBaseline and aminotransferase = 3xBaseline respectively. For total bilirubin values > 2xBaseline, values identified as > 2 ULN will be plotted differently than values ≤ 2 ULN.
      ii. Express total bilirubin in xULN on the y-axis and aminotransferase in xBaseline on the x-axis. Add horizontal and vertical lines corresponding to total bilirubin = 2 ULN and aminotransferase = 3xBaseline respectively.

Occurrence of aminotransferases (ALT and AST) increase will be presented separately according to the baseline level (normal or abnormal). The classes considered aminotransferase increase are:

1) > 3ULN and ≤ 5ULN, > 5ULN and ≤ 10ULN, > 10ULN if the baseline level of aminotransferase is normal (i.e. < ULN). Among all visits, the worst case will be retained.

2) > 3xBaseline and ≤ 5xBaseline, > 5xBaseline if the baseline level of aminotransferase is abnormal (i.e. ≥ ULN). Among all visits, the worst case will be retained.

Occurrence of CPK increase will be presented according to the baseline level (normal or abnormal). The classes considered for CPK increase are: > 3ULN and ⩽ 5ULN, > 5ULN. Among all visits, the worst case will be retained.

The increase in serum creatinine will be presented as:

1) Occurrence of at least one post baseline value > ULN (only on the subset of subjects with baseline value ≤ ULN)
2) Change from baseline in creatinine presented in classes according to KDIGO (Kidney Disease Improving Global Outcomes) AKI (Acute Kidney Injury) stages. Among all visits, the worst case will be retained:
    a. ≥ 1.5xBaseline and < 2.0xBaseline, or ≥ 0.3 mg/dL increase
    b. ≥ 2.0xBaseline and < 3.0xBaseline
    c. ≥ 3.0xBaseline or ≥ 4 mg/dL.

The relative decrease from baseline in eGFR (estimated glomerular filtration rate) will be presented in classes: ≥ 0 % and < 25 %, ≥ 25 % and < 50 %, ≥ 50 % and < 75 %, ≥ 75 %. Among all visits, the worst case will be retained.

In addition, the number and percentage of patients with markedly abnormal clinical laboratory values will be summarized for each parameter by time-point.

For each laboratory analyte, the baseline value will be defined as last scheduled or unscheduled value collected prior to the first dose of study treatment. Assessments carried out on day of first study treatment administration are considered to have taken place before the study treatment administration, if the corresponding times have not been recorded.

6.8.4 Vital Signs

Vital Signs to be summarized include:

- systolic blood pressure (mmHg);
- diastolic blood pressure (mmHg);
- pulse rate (bpm);
- weight (kg);
- waist circumference (cm);

Vital sign data recorded at each time-point (baseline, week 12, week 24, week 36, week 48, week 60, week 72, week 96, at every subsequent 24 weeks, at the end of the 72 week treatment period and at the end of the LTTP) and change from baseline to each post baseline time-point will be summarized by treatment group and overall as continuous data.

Note: waist circumference is not collected at week 12, week 36 or week 60, so will not be included in the summary tables at these time-points.

The baseline value will be defined as last scheduled or unscheduled value collected prior to the first dose of study treatment. Assessments carried out on day of first study treatment administration are considered to have taken place before the study treatment administration, if the corresponding times have not been recorded.

6.8.5 Electrocardiograms

An overall Investigator assessment of ECG will be provided (categories "normal", "abnormal, not clinically significant" and "abnormal, clinically significant").

The Investigator assessment will be listed and the number and percentage of patients within each assessment category will be tabulated by treatment group and overall and visit (baseline, week 36, week 72, week 120, at every subsequent 48 weeks, at the end of the 72 week treatment period and at the end of the LTTP).

Shifts from baseline (normal vs. abnormal, not clinically significant vs. abnormal, clinically significant) at the end of the 72 week treatment period and to the last post-baseline visit will be presented.

6.8.6 Physical Examination

Abnormalities identified from physical examination are recorded in the eCRF as Medical History or Adverse Events as appropriate and will be listed and summarized as such [See Sections 6.5.1 (Medical History) and 6.8.2 (Adverse Events)].

Physical examination results (normal/abnormal) and details of abnormalities will be listed for each subject.

6.8.7 Other Safety Variables

6.8.7.1 Other Safety Markers

Other safety markers to be summarized include NT-ProBNP, troponin-T, Homocysteine and cystatin C. NT-ProBNP will be log-transformed for the purpose of summaries and analysis.

Other safety marker values recorded at each time-point (baseline, week 24, week 48, week 60, week 72, week 96, at every subsequent 24 weeks, at the end of the 72 week treatment period and at the end of the LTTP) and change from baseline to each post baseline time-point will be summarized by treatment group and overall, and by endpoint, as continuous data.

Line plots of means with whiskers extending to the mean +/ SE over time using the same visits as for table summaries will be presented for homocysteine and cystatin C.

6.8.7.2 Drug Induced Liver Injury

As stated in the protocol, the DILI will be adjudicated by the Clinical Events Committee (CEC). The number and percentage of patients with adjudicated DILI events confirmed by the CEC will be summarized by treatment group.

6.8.7.3 France-specific procedures and markers

For patients enrolled at sites in France, additional summaries will be provided for the FSS. Occurrence of clinically significant abnormalities reported upon Ultrasonographic exam of bladder and urinary tract will be summarized at each time-point by treatment group and overall as categorical data.

Additionally, descriptive summaries of alpha-fetoprotein and osteocalcin recorded at each time point and change from baseline to each post-baseline time-point will be summarized by treatment group and overall as continuous data.

6.9 Interim Analysis

The interim analysis planned as per the protocol section 9.8.2 will not be performed.

7 CHANGES IN PLANNED ANALYSES

Following the review of the surrogate endpoint analysis (SEA) in 2020 and based upon limited efficacy results, GENFIT made the decision to prematurely terminate the trial. Consequently, the following planned analyses have been changed:

- The definition of the primary estimand was expanded to include the attribute of treatment comparison, in line with the final ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials, effective 30 July 2020.
- The final analysis will be performed at the time of the premature database lock, instead of at the time when 456 F2 and F3 experience a clinical event
- The scheduled interim analysis as described in section 9.8.2 of the protocol will not be performed.
- Primary study objective to evaluate the efficacy of elafibranor 120 mg versus placebo on clinical outcomes will not be performed. Instead, the clinical outcomes for the elafibranor 120 mg QD and placebo will be described and no significance testing will be performed.
- Secondary study objectives to assess histological changes after 72 weeks of treatment and at the end of the LTTP will not be performed
- Secondary study objectives on the following endpoints at week 72 and at the end of the LTTP will not be performed:
    - changes in liver enzymes and liver markers
    - changes in noninvasive markers of fibrosis and steatosis
    - changes in lipid parameters
    - variation in body weight
    - changes in insulin resistance and glucose homeostasis markers
    - changes in inflammatory markers
    - changes in cardiovascular risk profile as assessed by Framingham scores
    - changes in liver stiffness by FibroScan measurement
    - changes in quality of life (36-Item Short-Form Health Survey (SF-36) questionnaire)
- Secondary efficacy objective to assess the onset of the following endpoints will not be performed:
    - histological liver cirrhosis
    - death of any cause
    - any portal hypertension or cirrhosis related events

  - liver-related death events.
- Secondary efficacy objective to assess the onset of cardiovascular events will not be performed, instead, the onset of cardiovascular events will be described, and no significance testing will be performed
- The following analysis sets will no longer be defined and used for reporting:
  - Safety Set
  - Intent-to-treat Set
  - Per Protocol Set
  - Intent-to-Treat F1 Cohort Set
  - Safety F1 Cohort Set
- No sensitivity or subgroup summaries will be performed
- Due to the COVID-19 pandemic occurring concurrently with ongoing trial activities, the following additional measures have been added to describe the prevalence of the virus: SMQs will be used to classify each AE as COVID-19 or non-COVID-19 related;
  - Patient disposition will be separately presented for those patients with confirmed or suspected COVID-19;
  - Primary reasons for discontinuation will be supplemented with a breakdown of patients discontinuing for COVID-19 reasons;
  - Study completion will be separately presented for those patients with confirmed or suspected COVID-19;
  - An additional summary will be presented of all important deviations related to COVID-19;
  - Demographic characteristics will be additionally described for patients with confirmed or suspected COVID-19;
  - AE summaries to repeat general AE summary counts to include the number of patients and absolute number of COVID-19 related AEs in differing categories;
  - An additional summary will be presented for AEs related to COVID-19 by SOC and PT.
- Additional descriptive summaries of ultrasonography of bladder and urinary tract, alpha-fetoprotein and osteocalcin have been added for patients enrolled at sites in France.

8 DATA ISSUES

There are no data issues.

9 REFERENCES

- ICH. *Statistical Principles for Clinical Trials*, Guideline E9, 1998. Available at http://www.emea.eu.int/pdfs/human/ich/036396en.pdf
- ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. Available at https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles_en.pdf
- CPMP. *Points to Consider on Missing Data*. EMEA: London, 2001. Available at http://www.emea.eu.int/pdfs/human/ewp/177699EN.pdf
- CPMP. *Points to Consider on Adjustment for Baseline Covariates*. EMEA: London, 2003
- ICH. *ICH E3 Guideline: Structure and Content of Clinical Study Reports Questions & Answers*, 2012. Available at

http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E3/E3_QAs_R1_Step4.pdf

- KDIGO. *Kidney Disease: Improving Global Outcomes (KDIGO) clinical practice guideline for acute kidney injury. Kidney International Supplement*. 2012; 2: 19-36. Available at https://kdigo.org/wp-content/uploads/2016/10/KDIGO-2012-AKI-Guideline-English.pdf

10 APPENDICES

Appendix I - Schedule of Events

Table 3: STUDY GENERAL ASSESSMENT SCHEDULE

	Screening Period			First Treatment Period							Long-term Treatment Period		End Of Study Treatment
Visit	SV1	SV2 [4]	SPV 5	V1	V2	V3	V4	V5	V6	V7	Phone Visit PV1-PVn	V8-Vn	EOT [13]
Week	*-12 to -8*	*-12 to -4*	*-1* [5]	*0* [6]	*12* [6]	*24* [6]	*36* [6]	*48* [6]	*60* [6]	*72* [6]	*(every 24 weeks starting 12 weeks after V7)* [8]	*(every 24 weeks starting after V7)* [9]	*30 days after final study drug administration*
Permitted Margin				0	±1 week after V1	± 1 week after V1	± 1 week after V1	± 1 week after V1	± 1 week after V1	± 1 week after V1	± 2 weeks Compared to V7	± 2 weeks Compared to V7	± 1 week after last administration
Obtain informed consent	X												
Medical history / demographics	X												
Check inclusion / exclusion criteria	X			X [7]									
Adequate diet and lifestyle recommendations, including alcohol restrictions and smoking habits	X - - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - - →	
Confirmation of diet and lifestyle compliance, including alcohol restrictions and smoking habits				X	X	X	X	X	X	X	X	X	X
Physical examination	X			X	X	X	X	X	X	X		X	X
Vital signs & height [1] & weight measurement	X			X	X	X	X	X	X	X		X	X
Waist circumference	X			X		X		X		X		X	X

	Screening Period			First Treatment Period							Long-term Treatment Period		End Of Study Treatment
Visit	**SV1**	**SV2** [4]	**SPV** 5	**V1**	**V2**	**V3**	**V4**	**V5**	**V6**	**V7**	**Phone Visit PV1-PVn**	**V8-Vn**	**EOT** [13]
Week	*-12 to -8*	*-12 to -4*	*-1* **5**	*0* **6**	*12* **6**	*24* **6**	*36* **6**	*48* **6**	*60* **6**	*72* **6**	*(every 24 weeks starting 12 weeks after V7)* **8**	*(every 24 weeks starting after V7)* **9**	*30 days after final study drug administration*
Permitted Margin				**0**	**±1 week after V1**	**± 1 week after V1**	**± 1 week after V1**	**± 1 week after V1**	**± 1 week after V1**	**± 1 week after V1**	**± 2 weeks** Compared to V7	**± 2 weeks** Compared to V7	**± 1 week** after last administration
12-Lead ECG				X			X			X		X [10]	X
Lab evaluation (see Table 4)	X	X		X	X	X	X	X	X	X		X	X
Send sample for central histological evaluation of NASH diagnosis / change	X	X								X		X [11]	
Liver biopsy		X 4								X		X [11]	
Phone call to patient to confirm eligibility of histology criteria			X 5										
FibroScan 2				X						X		X	
Contact the patient prior to visit 3				X	X	X	X	X	X	X		X	X
Randomization				X									
IXRS registration	X			X	X	X	X	X	X	X	X	X	X
Review prior / concomitant medication	X			X	X	X	X	X	X	X	X	X	X
Quality of life assessment				X		X		X		X		X [12]	X

	Screening Period			First Treatment Period							Long-term Treatment Period		End Of Study Treatment
Visit	**SV1**	**SV2** [4]	**SPV** 5	**V1**	**V2**	**V3**	**V4**	**V5**	**V6**	**V7**	**Phone Visit PV1-PVn**	**V8-Vn**	**EOT** [13]
Week	*-12 to -8*	*-12 to -4*	*-1* 5	*0* 6	*12* 6	*24* 6	*36* 6	*48* 6	*60* 6	*72* 6	*(every 24 weeks starting 12 weeks after V7)* 8	*(every 24 weeks starting after V7)* 9	*30 days after final study drug administration*
Permitted Margin				**0**	**±1 week after V1**	**± 1 week after V1**	**± 1 week after V1**	**± 1 week after V1**	**± 1 week after V1**	**± 1 week after V1**	**± 2 weeks Compared to V7**	**± 2 weeks Compared to V7**	**± 1 week after last administration**
Adverse events	X	X		X	X	X	X	X	X	X	X	X	X
Data collection on clinical outcomes					X	X	X	X	X	X	X	X	X
Study placebo or drug dispensation				X	X	X	X	X	X	X		X	
Drug accountability					X	X	X	X	X	X	X	X	X

Abbreviations: ECG = electrocardiogram; EOT = end of treatment; IXRS = Interactive voice/web Response System; LTTP = Long-term Treatment Period; NASH = nonalcoholic steatohepatitis; PV = phone visit; QOL = quality of life; SV = Screening visit; V = visit

1. Height is measured only at visit SV1.
2. Where possible FibroScan must be done at the day of visit. Otherwise, it can be performed within 7 days around the visit date.
3. During the study, the patient should be contacted at least 1 week before the next visit as a reminder on procedures and IP return.
4. This visit only occurs if no historical biopsy within 6 months before the Screening Visit is available. A screening liver biopsy and slides shipment to the central anatomopathologist must be performed at least 4 weeks before Randomization (in order to obtain the results in time). Coagulation (platelet count and PT [INR]) should be checked locally prior to this liver biopsy (according to local medical standards in each hospital).
5. Screening Phone Visit. Telephone contact for all patients at least 1 week before V1. Patients should be contacted regarding eligibility confirmation within 1 week prior to Randomization Visit V1. In case of ineligibility, the patient should be contacted as soon as possible.
6. The maximum time period between visits in the First Treatment Period is to be 96 days due to the study drug supply provided to the patient.
7. Check of all inclusion/exclusion criteria, including biological and histological criteria assessed at SV1 and SV2.

8. Phone visits every 24 weeks starting 12 weeks after V7 for safety, data collection on clinical outcomes, and IP compliance control. If any information during this phone contact requires a visit, the Investigator may decide to perform an unscheduled visit. Phone visits may also be performed at the same frequency for the follow-up of patients having permanently discontinued study drug but remaining in the study (Same information collected except IP compliance control).
9. The maximum time period between visits in the Long-term Treatment Period (LTTP) is to be 192 days due to the study drug supply provided to the patient.
10. In the LTTP the first ECG will be performed at V9 and then every 48.
11. Liver biopsy will be performed after approximately 4 years of treatment (V13) and in the case of suspicion of cirrhosis (based on FibroScan and/or clinical or biological assessment). In the case that the suspicion of cirrhosis is not confirmed by liver biopsy the patient shall remain on the study and a liver biopsy will be performed after approximately 4 years of treatment (V13, unless a biopsy has already been performed within the year). Blood sampling (coagulation tests; see) are to be performed locally before the biopsy.
12. QOL assessment questionnaire to be completed at 24 (V8), 48 (V9), and 96 (V11) weeks in the LTTP (following approximately 96, 120, and 168 weeks of treatment, respectively), and every 48 weeks thereafter.
13. EOT Visit to be performed 30 days after final study drug administration at the end of study or for any premature discontinuation (permanent study drug discontinuation or trial discontinuation).

Table 4: STUDY BIOLOGICAL ASSESSMENT SCHEDULE

Visit	Screening Period		First Treatment Period							Long-term Treatment Period	End of Study Treatment
	SV1 SB1 6	SB2 7	V1 B1	V2 B2	V3 B3	V4 B4	V5 B5	V6 B6	V7 B7	V8 to Vn B8 to Bn	EOT
Week	*-12 to -8*	*Prior to -4*	*0*	*12*	*24*	*36*	*48*	*60*	*72*	*Every 24 weeks*	*30 days after final study drug administration*
Hematology *Hemoglobin, hematocrit, RBC, WBC, differential count, platelet count, reticulocytes count, and PT (INR)*	X		X	X	X	X	X	X	X	X	X
Coagulation - local lab testing prior to liver biopsy *Platelet count, PT (INR)* 1		X							X	X 1	
Serology *HIV ab I/II, HBsAg, and HCV Ab (positive HCV RNA in case HCV Ab >0 or known cured hepatitis C infection* 2*)*	X										
Screening Visit 1 - chemistry panel *HbA1c*2*, fasting plasma glucose, insulin (fasting), HOMA-IR creatinine, eGFR, GGT, AST, ALT, CPK*2*, alkaline phosphatase, total and conjugated bilirubin, sodium, TG, and MELD score*	X										
V1 to Vn total chemistry panel *HbA1c, fasting plasma glucose, creatinine, eGFR, GGT, AST, ALT, CPK, alkaline phosphatase, total proteins, albumin, electrolytes (sodium, potassium, chloride, calcium), uric acid, urea (BUN), total and conjugated bilirubin, hsCRP, total cholesterol, nonHDL-C, HDL-C, TG, calculated VLDL-C, ApoAI, ApoB, calculated LDL-C, and MELD score*			X 6	X	X	X	X	X	X	X	X

Visit	Screening Period		First Treatment Period							Long-term Treatment Period	End of Study Treatment
	SV1 SB1 6	SB2 7	V1 B1	V2 B2	V3 B3	V4 B4	V5 B5	V6 B6	V7 B7	V8 to Vn B8 to Bn	EOT
Week	*-12 to -8*	*Prior to -4*	*0*	*12*	*24*	*36*	*48*	*60*	*72*	*Every 24 weeks*	*30 days after final study drug administration*
Urinalysis *albumin, creatinine, ACR, and microscopic analysis* *α1 microglobulin*, β-NAG,* N-Gal*, IL-18*, KIM-1*,*			X	X	X	X	X	X	X	X	X
Urinalysis (dipstick) *Specific gravity, pH, protein, glucose, ketones, bilirubin, urobilinogen, blood, nitrite, and leukocytes*	X		X	X	X	X	X	X	X	X	X
Urinary pregnancy tests 3	X		X	X	X	X	X	X	X	X	X
Inflammatory markers *Fibrinogen, and haptoglobin*			X		X		X		X	X	X
Other Liver markers *CK18 (M65 & M30), adiponectin, ferritin, FGF19 & FGF21, alpha2 macroglobulin, hyaluronic acid, PIIINP, TIMP-1, and CHI3L1* 4			* 4		*		*		* 4	* 4	*
Calculated fibrosis & steatosis index *Fibrotest, ELF, NAFLD Fibrosis score, Steatotest, FLI, Fibrometre S, and FIB-4*			*		*		*		*	*	*
Other safety markers *Homocysteine, NT-ProBNP, troponin-T, and cystatin C*			*		*		*		*	*	*
Special glycemic and other lipid parameters *Insulin(fasting), HOMA-IR, Fructosamine, C-peptide, FFA, small dense LDL, ApoAII, Apo CIII, and Apo E*			*		*		*		*	*	*

Visit	Screening Period		First Treatment Period							Long-term Treatment Period	End of Study Treatment
	SV1 SB1 6	SB2 7	V1 B1	V2 B2	V3 B3	V4 B4	V5 B5	V6 B6	V7 B7	V8 to Vn B8 to Bn	EOT
Week	*-12 to -8*	*Prior to -4*	*0*	*12*	*24*	*36*	*48*	*60*	*72*	*Every 24 weeks*	*30 days after final study drug administration*
Sampling for additional parameters *Whole blood 5, plasma, and serum bank*	* 5		*	*	*	*	*	*	*	*	*

X = results available within 2 working days (routine analysis) * = batch analysis

Abbreviations Ab = antibody; ACR = albumin–creatinine ratio; Ag = antigen; ALT = alanine aminotransferase; Apo = apolipoprotein; AST = aspartate aminotransferase; β-NAG = N-acetyl-β-D-glucosaminidase; BUN = blood urea nitrogen; B = biological assessment Visit; CHI3L1 = chitinase-3-like protein 1; CK18 = cytokeratin 18; CPK = creatine phosphokinase; eGFR = estimated glomerular filtration rate; ELF = enhanced liver fibrosis; EOT = end of study treatment; FFA = free fatty acid; FGF = fibroblast growth factor; FIB-4 = fibrosis 4 score; FLI = fatty liver index; GGT = gamma-glutamyl transferase; HbA1c = hemoglobin A1c; HBsAg = hepatitis B surface antigen; HCV = hepatitis C virus; HDL-C = high density lipoprotein-C; HIV = human immunodeficiency virus; HOMA-IR = homeostasis model assessment of insulin resistance; hsCRP = high-sensitivity C-reactive protein; IL-18 = interleukin 18; INR = international normalized ratio; KIM-1 = kidney injury molecule-1; LDL-c = low density lipoprotein-C; MDRD = modification of diet in renal disease; MELD = model end stage liver disease; NAFLD = nonalcoholic fatty liver disease; N-Gal = neutrophil gelatinase-associated lipocalin; NT-ProBNP = N-terminal of the prohormone brain natriuretic peptide; PIIINP = type III procollagen peptide; PT = prothrombin time; TIMP-1 = tissue inhibitors of metalloproteinases 1; RBC = red blood cell; SB = Screening biological assessment Visit; SV = Screening Visit; TG = triglyceride; VLDL-C = very low density lipoprotein-C; V = Visit; WBC = white blood cell.

1. Coagulation (platelet count and PT [INR]) should be checked prior to any liver biopsy (according to local medical standards in each hospital). To be done through a local laboratory. Liver biopsy will be performed after 72 weeks (V7) and after approximately 4 years of treatment (V13) and in the case of suspicion of cirrhosis (based on FibroScan and/or clinical or biological assessment). In the case that the suspicion of cirrhosis is not confirmed by liver biopsy the patient shall remain on the study and a liver biopsy will be performed after approximately 4 years of treatment ([V13] unless a biopsy has already been performed within the past year).
2. Upon receipt of the results of the biological assessment performed at SV1, retesting or additional testing may be needed during the Screening Period:
   - CPK can be repeated prior to Randomization (V1) within 1 to 2 weeks after initial test.
   - HbA1c can be repeated prior to Randomization (V1), at the latest 2 weeks prior to planned Randomization.
   - HCV RNA can be tested, at SV1 in case of known cured hepatitis c infection, or in case of positive HCV Ab at SV1, at a retest screening visit at the latest 2 weeks prior to the planned Randomization (V1).
3. Dipstick at site for WOCBP only. In addition, home pregnancy tests are to be performed by WOCBP every 4 weeks from V1.
4. CHI3L1 to be tested only at V1, V7, and at the time of 4 years biopsy (V13).

5. Whole blood sample will be only taken at SV1 while plasma and serum samples are to be taken at every visit **ONLY** for patients who have signed the pharmacogenomic and biomarker ICF.
6. Visits SV1 and V1 should be scheduled at least 8 weeks apart in order to have 2 consecutive baseline values of AST, ALT, total bilirubin, and INR for DILI adjudication.
7. SB2, additional visit in the Screening Period if required for coagulation prior to liver biopsy.

Appendix II - Table, Figure and Listing Shells

The table, figure and listing shells and corresponding Table of Contents are in a separate file.