Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

STATISTICAL ANALYSIS PLAN
(Double-Blind Period)

Protocol GFT505B-319-1

A Double-blind, Randomized, Placebo-Controlled Study and Open-label Long Term Extension to evaluate efficacy and safety of Elafibranor 80 mg in Patients with Primary Biliary Cholangitis with Inadequate Response or Intolerance to Ursodeoxycholic Acid

Protocol Number: **(Version Date)**	GFT505B-319-1, v4.0, IND number: 132202 20-September-2022
Name of Test Drug:	Elafibranor
Phase:	3
Methodology:	A Double-blind, Randomized, Placebo-Controlled Study and Open-label Long Term Extension to evaluate efficacy and safety of Elafibranor 80 mg in Patients with Primary Biliary Cholangitis with Inadequate Response or Intolerance to Ursodeoxycholic Acid.
Sponsor:	GENFIT Parc Eurasanté 885, Avenue Eugène Avinée 59120 LOOS, France, Phone: PPD
Document Date:	10-May-2023
Document Version:	4.0

CCI [redacted]

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

SIGNATURE PAGE

Protocol Title:	A Double-blind, Randomized, Placebo-Controlled Study and Open-label Long Term Extension to evaluate efficacy and safety of Elafibranor 80 mg in Patients with Primary Biliary Cholangitis with Inadequate Response or Intolerance to Ursodeoxycholic Acid.
Sponsor:	GENFIT Parc Eurasanté 885, Avenue Eugène Avinée 59120 LOOS, France, Phone: PPD
Protocol Number:	GFT505B-319-1, IND number: 132202
Document Date/Version:	10-May-2023/v4.0

Cytel, Inc. Author:
PPD
Principal Biostatisticians

Signature:
Date:

Signature:
Email: PPD

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

Sponsor Approval

By signing this document, I acknowledge that I have read the document and approve of the planned statistical analyses described herein. I agree that the planned statistical analyses are appropriate for this study, are in accordance with the study objectives, and are consistent with the statistical methodology described in the protocol, clinical development plan, and all applicable regulatory guidance's and guidelines.

I have discussed any questions I have regarding the contents of this document with the biostatistical author.

I also understand that any subsequent changes to the planned statistical analyses, as described herein, may have a regulatory impact and/or result in timeline adjustments. All changes to the planned analyses will be described in the clinical study report (CSR).

Sponsor Signature : GENFIT 885, Avenue Eugène Avinée, 59120 Loos, France	Date : _____
PPD, Statistician	Date : **Signature:** PPD **Email:**
PPD, VP Biometrics	Date : **Signature:** PPD **Email:**
PPD, CMO	Date : **Signature:** PPD **Email:**

CCI

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

MODIFICATION HISTORY

Unique Identifier for SAP Version	Date of SAP Version	Author	Changes from the Previous Version
			Initial issuance of document
V1.0	28-July-2020	PPD	First signed version. Based on protocol version 0.7, 15-July- 2020.
V2.0	16-September-2021	PPD	Addressing FDA's feedback; adding the optimal weighting scheme as proposed in Song et al. (2013).
V3.0	27-October-2022	PPD	Updates based on protocol version 4.0, 20Sept22; Addressing FDA's feedback in the Informational Request and Writing Response Only; Change analysis population for the 2nd key secondary endpoint related to pruritus; addition of 3rd key secondary endpoint (change in NRS through week 24)
V4.0	10-May-2023	PPD	Update based on FDA's feedback in the Informational Request and Writing Response Only on February 17th, 2023.

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

TABLE OF CONTENT

1. INTRODUCTION AND OBJECTIVES OF ANALYSIS 15
   1.1 Introduction 15
   1.2 Objectives of Statistical Analysis 15
2. STUDY DESIGN 16
   2.1 Synopsis of Study Design 16
   2.2 Screening period 16
   2.3 Baseline period (V1) 17
   2.4 Double-Blind period (Week 0 up to Week 104) 18
   2.5 Long-Term extension 19
   2.6 Clinical event committee (CEC) 19
   2.7 Optional Visits 20
      2.7.1 Retesting and/or additional screening visits 20
      2.7.2 Unscheduled visits 21
   2.8 Exploratory/Ancillary Sub-study 21
   2.9 Randomization Methodology 21
   2.10 Stopping Rules and Unblinding 22
   2.11 Study Procedures 23
   2.12 Efficacy, Pharmacokinetic, and Safety Variables 34
      2.12.1 Efficacy Variables 34
      2.12.2 Pharmacokinetic Variables 36
      2.12.3 Safety Variables 36
      2.12.4 Exploratory Variables (related to histological assessment) 37
3. SUBJECT POPULATIONS 38
   3.1 Population Definitions 38
   3.2 Protocol Deviations 40
4. STATISTICAL METHODS 41
   4.1 Sample Size Justification 41
   4.2 General Statistical Methods and Data Handling 41
      4.2.1 General Methods 41
      4.2.2 Data Conventions 42

CCI [redacted]

4.2.3 Computing Environment 44
4.2.4 Methods of Pooling Data 44
4.2.5 Adjustments for Covariates 44
4.2.6 Multiple Comparisons/Multiplicity 45
4.2.7 Subpopulations 45
4.2.8 Withdrawals, Dropouts, Loss to Follow-up 46
4.2.9 Missing, Unused, and Spurious Data 46
4.2.10 Visit Windows 50
4.2.11 Data Derivation 51
4.3 Interim Analyses 52
4.4 Subject Disposition 53
4.5 Demographic, Screening and Baseline Characteristics 53
4.5.1 Demographics 54
4.5.2 Screening Characteristics 54
4.5.3 Baseline Disease Characteristics 55
4.5.4 Prior Medications 56
4.5.5 Prior Procedures 56
4.6 Efficacy Evaluation 56
4.6.1 Estimand Framework for Primary and Key Secondary Endpoints 56
4.6.2 Primary Efficacy endpoint analysis 62
4.6.2.1 Main Analysis 62
4.6.2.2 Sensitivity analysis to the statistical model 64
4.6.2.3 Supplementary analyses 64
4.6.2.3.1 Composite strategy with missing data as non-responders 64
4.6.2.3.2 Hypothetical strategy 64
4.6.2.3.3 Treatment policy strategy 72
4.6.2.3.4 COVID-19 74
4.6.3 First Key Secondary Efficacy Endpoint 74
4.6.4 Second Key Secondary Efficacy Endpoint 75
4.6.4.1 Main Analysis 75
4.6.4.2 Supplementary analysis 76

4.6.4.2.1 Tipping point analysis 76
4.6.4.2.2 Other supplementary analysis 78
4.6.5 Third Key Secondary Efficacy Endpoint 80
4.6.6 Subgroup Analysis 81
4.6.7 Other secondary endpoints 82
4.6.7.1 Change from baseline in ALP at 4, 13, 26, 39 and 52 weeks (continuous) 82
4.6.7.2 ALP response defined as 10%, 20% and 40% ALP reduction from baseline at week 52 (binary) 83
4.6.7.3 Response to treatment according to ALP < 1.5x ULN, ALP decrease from baseline ≥ 40% and TB ≤ ULN at week 52 (binary) 83
4.6.7.4 Response to treatment according to ALP < 3x ULN, AST <2x ULN and TB ≤ 1 mg/dL (Paris I) at week 52 (binary) 83
4.6.7.5 Response to treatment according to ALP ≤ 1.5x ULN, AST ≤ 1.5x ULN and TB ≤ 1mg/dL (Paris II) at week 52 (binary) 84
4.6.7.6 Response to treatment according to TB decrease of 15% change from baseline at week 52 (binary) 84
4.6.7.7 Response to treatment according to normalization of abnormal TB (TB ≤ ULN) and/or ALB (ALB ≥ LLN) (Rotterdam) at week 52 (binary) 84
4.6.7.8 Response to treatment according to TB ≤ 0.6x ULN at week 52 (binary) 84
4.6.7.9 Response to treatment according to ALP ≤ 1.67x ULN and TB ≤ 1mg/dL (Momah/Lindor criterion, 2011) at week 52 (binary) 84
4.6.7.10 No worsening of TB defined as TB ≤ ULN at week 52 or no increase from baseline of more than 0.1x ULN at week 52 (binary) 84
4.6.7.11 Complete biochemical response defined as normal ALP, TB, AST, ALT, albumin, and INR at week 52 (binary) 85
4.6.7.12 UK-PBC score at week 52 (continuous) 86
4.6.7.13 GLOBE PBC score at week 52 (continuous) 86
4.6.7.14 Response to treatment based on the proportion of patients who normalized bilirubin (TB ≤ ULN) at week 52 (binary) 87
4.6.7.15 Response to treatment based on the proportion of patients who normalized albumin (ALB ≥ LLN) at week 52 (binary) 87
4.6.7.16 Change from baseline to week 52 in hepatobiliary injury and liver function as measured by AST, ALT, GGT, 5' NT, total and conjugated bilirubin, albumin, INR and ALP fractionated (hepatic) (continuous) 88
4.6.7.17 Change from baseline to week 52 in biomarkers of inflammation as measured by hsCRP, CCI

fibrinogen, haptoglobin and TNF-α (continuous) 89

4.6.7.18 Change from baseline to week 52 in immune response as measured by IgG and IgM (continuous) 89

4.6.7.19 Change from baseline to week 52 in biomarkers, non-invasive measures of hepatic fibrosis as measured by ELF test (HA, PIINP, TIMP-1), PAI-1, TGF-β, CK-18 (M65 and M30), Pro-C3 and liver stiffness measured by Transient Elastography (continuous) 89

4.6.7.20 Change from baseline to week 52 in lipid parameters as measured by TC, LDL-C, HDL-C, calculated VLDL-C and TG (continuous) 89

4.6.7.21 Change from baseline to week 52 in Fasting Plasma Glucose (continuous) 89

4.6.7.22 Change from baseline to week 52 in bile acids and biomarkers of bile acid synthesis as measured by bile acids, C4 and FGF-19 (continuous) 90

4.6.7.23 No worsening of pruritus based on PBC Worst Itch NRS through week 52 and through week 24 (binary) 90

4.6.7.24 Proportion of responders in PBC Worst Itch NRS score according to clinically meaningful change; at least 30% reduction; and one point, two points or three points decrease in score from baseline through week 52 and through week 24 in patients with baseline PBC Worst Itch NRS score ≥ 4 90

4.6.7.25 Proportion of participants achieving sustained improvement in PBC Worst Itch NRS (binary) 91

4.6.7.26 PGIC and PGIS 91

4.6.7.27 Change from baseline to week 52 in 5-D Itch (continuous) 92

4.6.7.28 Change from baseline to week 52 in PROMIS Fatigue Short Form 7a (continuous) 92

4.6.7.29 Change from baseline to week 52 in Epworth Sleepiness Scale (continuous) 93

4.6.7.30 Change from baseline to week 52 in PBC-40 (continuous) 93

4.6.7.31 Change from Baseline to week 52 in health utility as measured by the EQ-5D-5L (continuous) 94

4.6.7.32 Onset of clinical outcomes described as a composite endpoint of MELD-Na; Liver transplant; Uncontrolled ascites requiring treatment; Hospitalization for Variceal bleed, for Hepatic encephalopathy and for spontaneous bacterial peritonitis; death until week 52 (Binary) 94

4.6.7.33 Change from baseline to week 52 in serum markers of bone turnover and in bone mineral density (hip and lumbar) assessed by DEXA scanning (continuous) 95

4.7 Pharmacokinetic Evaluations 96

4.8 Safety Analyses 96

4.8.1 Extent of Exposure and Compliance 96

4.8.1.1 Extent of Exposure 96

4.8.1.2 Compliance 97
4.8.2 Adverse Events 97
4.8.3 Adjudicated events 102
4.8.4 Laboratory Data 103
4.8.5 Transient elastography 106
4.8.6 Vital Signs and Physical Examinations 107
4.8.7 Electrocardiogram 107
4.8.8 Concomitant Medications 108
4.8.9 Concomitant Procedures 108
4.9 Exploratory analyses (related to histological assessments) 109
5. CHANGES TO PLANNED ANALYSES 111
6. REFERENCES 112
7. APPENDIX A 114
8. APPENDIX B 115
9. APPENDIX C 116
10. APPENDIX D 119
11. APPENDIX E 124
12. APPENDIX F 126
13. STATISTICAL OUTPUTS TO BE GENERATED 136

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

LIST OF IN-TEXT TABLES

Table 1: Schedule of Assessments 23
Table 2: Schedule of Biological Assessments 28
Table 3: Schedule of Patient Reported Outcomes Questionnaires 32
Table 4: Missing Severity and Definition of Treatment-emergent 50
Table 5: Evaluation Intervals for Efficacy and Safety Analysis 50
Table 6: Histological Endpoints 109
Table 7: PROMIS Fatigue short form 7a Conversion table 116
Table 8: The six domains and 40 questions of the PBC-40 117
Table 9: Laboratory Test Toxicity Grades 119
Table 10: Classification of worst value for laboratory parameters of interest 124
Table 11: List of PT defining Hepatic Failure 126
Table 12: List of PT defining Hepatic Injury 128

LIST OF IN-TEXT FIGURES

Figure 1: Study Design Overview 20

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

ABBREVIATIONS

Abbreviation	Definition
ACR	Albumin to Creatinine Ratio
AE	Adverse event
AESI	Adverse event of specific interest
AFP	Alfa-fetoprotein
AKI	Acute kidney injury
ALB	Albumin
ALP	Alkaline phosphatase
ALT	Alanine transaminase
ANCOVA	Analysis of covariance
AST	Aspartate transaminase
AT	Aminotransferase
BDRM	Blind Data Review Meeting
CEC	Clinical event committee
CK-18	Cytokeratin-18
CI	Confidence Interval
CMH	Cochran-Mantel-Haenszel
CMO	Chief Medical Officer
CPK	creatine phosphokinase
CSR	Clinical study report
C4	Serum 7 α-hydroxy-4-cholesten-3-one
DB	Double-blind
DEXA	Dual-Energy X-ray Absorptiometry
DILI	Drug-induced Liver injury
DSMB	Data Safety Monitoring Board
5D-Itch	5-D Pruritus Scale
ECG	12-lead Electrocardiogram
eCRF	electronic Case Report Form
eDiary	Electronic diary

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

Abbreviation	Definition
eDISH	Evaluation of Drug-Induced Serious Hepatotoxicity
eGFR	estimated glomerular filtration rate
ELF	Enhanced Liver Fibrosis
EODB	End of the Double-Blind period
EOS	End of Study
EOT	End of treatment
ESS	Epworth Sleepiness Scale
FGF-19	Fibroblast growth factor 19
FPG	Fasting plasma glucose
GCA	Glycocholic acid
GGT	Gamma-glutamyltransferase
HA	Hyaluronic acid
Anti- HAV IgM	Hepatitis A Antibody test
HBsAg	Surface antigen of the hepatitis B virus (Australia antigen)
HCV Ab	Hepatitis C antibody
HCV RNA	Qualitative Detection of Hepatitis C Virus RNA
HDL-C	High-Density Lipoprotein Cholesterol
HIV Ab I/II	Rapid HIV antibody test
HRQoL	Health-related Quality of Life
hsCRP	high-sensitivity C-Reactive Protein
ICE	Intercurrent event
ICH	International Conference on Harmonization
ICF	Informed Consent Form
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IL-6	Interleukin 6
INR	International normalized ratio
ITT	Intent-to-treat
IxRS	interactive voice/web response system
KDIGO	Kidney disease improving global outcomes
LDL-C	Low-Density Lipoprotein Cholesterol

CCI [redacted]

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

Abbreviation	Definition
LLN	Lower limit of normal
LSM	Least Squares Means
LTE	long term extension
LVDB	Last Visit Double-Blind
M1, M2, M3	First Measure, Second Measure, Third Measure
M30	measures caspase-cleaved CK18 produced during apoptosis
M65	measures the levels of both caspase-cleaved and intact CK18, total cytokeratin
MAR	Missing At Random
MedDRA	Medical Dictionary for Regulatory Activities
MELD-Na	model end stage liver disease-Sodium
MMRM	Mixed effect Model for Repeated Measures
N/A	Not applicable
5' NT	5'-nucleotidase
OCA	Ocaliva (obeticholic acid)
PAI-1	Plasminogen activator inhibitor-1
PBC	Primary Biliary Cholangitis
PBC-40	Quality of Life for Primary Biliary Cirrhosis
PBI	placebo-based multiple imputation
PDP	Protocol Deviation Plan
PGIC	Patient Global Impression of Change
PGIS	patient global impression-severity
PIINP	Procollagen III amino terminal peptide
PK	Pharmacokinetic
PP	Per Protocol
PRO	Patient reported outcomes
Pro-C3	Plasma collagen type III
PCSA	Potentially clinically significant abnormality
Pt	prothrombin time
PT	Preferred Term
Q1, Q3	First Quartile, Third Quartile
QTc	corrected QT

CCI [redacted]

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

Abbreviation	Definition
RIM	Risk and Issue Management System
SAE	Serious adverse event
SAP	Statistical analysis plan
SOC	System Organ Class
SV	Screening visit
TC	Total cholesterol
TB	Total Bilirubin
TE	Transient elastography
TEAE	Treatment-emergent adverse event
TIMP-1	Tissue inhibitor of metalloproteinase 1
TG	Triglycerides
TGF-β	Transforming growth factor beta
TNF-α	tumor necrosis factor alpha
UDCA	Ursodeoxycholic acid
ULN	Upper limit of normal
US	Ultrasound exam
VLDL-C	Very-low-density lipoprotein Cholesterol
V1	Randomization/baseline visit
V2-V6	Visit2-Visit6
WBC	White cell blood count

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

1. INTRODUCTION AND OBJECTIVES OF ANALYSIS

1.1 Introduction

This document describes the plan for the statistical analyses and reporting of the double-blind (DB) period of the study GFT505B-319-1, which is a prospective, randomized, double-blind, placebo- controlled phase 3 study to demonstrate the efficacy of Elafibranor 80 mg versus Placebo in patients with Primary Biliary Cholangitis (PBC) and inadequate response or intolerance to ursodeoxycholic acid (UDCA). Results of the analyses described in this Statistical Analysis Plan (SAP) will be included in the Clinical Study Report (CSR).

Additionally, the planned analyses identified in this SAP will be included in regulatory submissions or future manuscripts. Any post-hoc or unplanned analyses related to the DB period and performed to provide results for inclusion in the CSR but not identified in this prospective SAP will be clearly identified in the CSR.

The information regarding the analyses of the open label long term extension (LTE) period is not within the scope of this SAP and will be described in a separate SAP.

1.2 Objectives of Statistical Analysis

The primary objective of the study GFT505B-319-1, as stated in the protocol is to evaluate the effect of Elafibranor (80 mg/day) on cholestasis over 52 weeks of treatment compared to placebo.

The key secondary objectives of the study are to evaluate the effect of Elafibranor (80 mg/day) over 52 weeks of treatment compared to placebo based on ALP normalization, and through 52 weeks and through 24 weeks on pruritus.

This SAP is designed to outline the methods to be used in the analysis of study data in order to address the study objectives. Populations for analysis, data handling rules, statistical methods, and formats for data presentation are provided in this SAP. The statistical analyses and summary tabulations described in this SAP will provide the basis for the results sections of the CSR for this study.

This SAP will also outline any differences in the currently planned analytical objectives relative to those planned in the study protocol.

2. STUDY DESIGN

2.1 Synopsis of Study Design

This is a phase 3 DB, randomized, placebo-controlled study evaluating Elafibranor 80 mg once daily versus placebo in patients with PBC and inadequate response or intolerance to UDCA.

2.2 Screening period

The study starts with a screening period, which must include at least one screening visit (SV) which should be performed between 2 to 12 weeks prior to randomization. At the SV1, preliminary eligibility criteria will be reviewed.

For the purpose of establishing relevant baseline chemistries for suspected drug-induced liver injury (DILI), repeated measures of aspartate transaminase (AST), alanine transaminase (ALT) and total bilirubin (TB) will be collected (see section 4.8.4) ONLY if the first measure (M1) collected at SV1 is > upper limit of normal (ULN).

M1 and another value, either a historical value (M0), collected at least 4 weeks and up to 12 weeks apart before SV1, or a second measure (M2) will be collected at SV2 (4 to 6 weeks after SV1). This applies to only the analyte above ULN.

- If variability between M1 and either M0 or M2 is ≤ 40% the patient is eligible
- If variability between M1 and either M0 or M2 is > 40% a third measure (M3) will be collected at SV3 (4 to 6 weeks after SV1 (if M0 compared) or after SV2 (if M2 compared)):
    - If variability between M1 and M3 is ≤ 40% the patient is eligible
    - If variability between M1 and M3 is > 40% the patient is ineligible for the study.

To assess the alkaline phosphatase (ALP) eligibility criterion, two ALP values will be required. One value is the M1 collected at SV1, and the other value is M0 collected at least 4 weeks and up to 12 weeks apart before SV1 or M2 collected at SV2 (4 to 6 weeks after SV1).

- If M1 is < 1.67x ULN, the patient will be excluded, and no further assessment will be performed.
- If M1 is ≥ 1.67x ULN, and the mean of M1 and either M0 or M2 is ≥ 1.67x ULN, and variability is ≤ 40%, the patient is eligible
- If M1 is ≥ 1.67x ULN, and the mean of M1 and either M0 or M2 is < 1.67x ULN or variability is > 40%, M3 collected at SV3 (4 to 6 weeks after SV1 (if M0 compared) or after SV2 (if M2 compared)) will be required:
    - If the mean of M1 and M3 is ≥ 1.67x ULN and variability is ≤ 40%, the patient is eligible
    - If the mean of M1 and M3 is < 1.67x ULN or variability is > 40%, the patient is ineligible for the study

NOTE: M2 and M3 values for AST, ALT, TB and ALP can be obtained locally. Comparison between M1 and local value(s) will be done on normalized values according to ULN.

The laboratory value used for the first stratification factor (see section 2.9) will be the ALP value and TB value provided by the central lab at SV1.

Local laboratory assessments are allowed for repeated assessment of ALT, AST, TB, ALP, and CPK during the screening period as well as for any required retest for liver function monitoring. Assessments of urinary myoglobin are to be done locally only in case of clinically significant CPK elevation.

At the screening period, the second stratification factor will be the PBC Worst Itch NRS score. PBC Worst Itch NRS score will be recorded on an electronic diary (eDiary) daily each evening; the PBC Worst Itch NRS score will be averaged over the 14 days preceding Baseline (V1: Day 1) to determine stratification for randomization. The PBC Worst Itch Score will be computed as an average of non-missing data falling in the 14-day window. At least 4 values of the PBC Worst Itch NRS score from each of the two 7-day intervals in the 14 days prior to randomization visit are required for randomizing the patients. If this number is not achieved, the screening period may be extended in order to obtain the expected number of NRS values.

At Screening the availability of historic liver biopsy samples may be determined to support the diagnosis of PBC. Liver biopsy is one of the three criteria to help identify PBC patients, along with two non-biopsy criteria. A minimum of two out of three criteria must be met in order to meet eligibility criteria (see protocol, inclusion criterion 3).

As a first step, patients will be asked if they agree to participate in the study and sign the Informed Consent Form (ICF). Each patient who has signed the ICF will perform procedures listed in Table 1. All inclusion and exclusion criteria will be reviewed. If eligibility is confirmed, the site will contact the patient to confirm the patient's willingness to continue in the study. In case of ineligibility, the patient should be contacted as soon as possible.

During the screening period, the patient will not receive study drug.

2.3 Baseline period (V1)

As per protocol, evaluations performed at the Study Day 1 visit (V1) should be done before drug administration and will be used for baseline values. Thus, baseline is defined as the last non-missing assessment the day of or prior to study drug start. If the assessment for ALP/AST/ALT/TB is not done at the baseline visit, a previous assessment by the central lab during the screening period closest to V1 may be used.
Note that the baseline of laboratory parameters assessed by central laboratory will be the last non-missing central assessment the day of or prior to study drug start.
Note that the baseline PBC Worst Itch NRS average will be calculated based on available data within 14 days prior the day of randomization.

2.4 Double-Blind period (Week 0 up to Week 104)

Patients who satisfy all eligibility criteria and confirm participation, will be randomized (see section 2.9) in a 2:1 ratio to one of two treatment groups:

- Elafibranor 80mg
- Placebo

In this DB period, study medication will be administered orally, once daily for up to 24 months. For any patient, the DB period will last until the last patient completes Visit 6 (W52) or until a maximum of 104 weeks, whichever happens first.

On the first day of the DB period the patient will receive the study drug after all baseline assessments are completed. During the entire double-blind period, the laboratory assessments (at every visit) will be done before study drug intake.

When applicable, patients should continue their pre-study dose of UDCA throughout the study participation. During the DB period the patients will return to the site for visits every 13 weeks (±2 weeks) from the Randomization Visit (V1) to Visit 6 except for the Visit 2 (4 weeks after V1), and if applicable every 26 weeks from Visit 6 to Visit 8 (W104). When the last V5 is performed, any patient between Visit 6 and Visit 8 should perform a last visit DB (LVDB) within the next 13 weeks. For patients who have not yet had V6 at the time the last patient completes his/her V5, LVDB will be scheduled at the latest 13 weeks after the date on which V5 was completed for the last patient. For those patients, LVDB will coincide with V6, and patients will complete V6 and associated procedures.

Patients who have not yet had V6 completed at the time the last patient completes his/her V5, will complete LVDB at the next scheduled visit (either V7 or V8) within the next 13 weeks. LVDB will replace V7 or V8. The same procedures as for V8 will be performed for LVDB (except US exam).

Between V6 and V8, safety contacts will be performed by phone call between on-site visits, every 26 weeks, to collect information related to AEs, concomitant medications, pregnancy, and study drug compliance. The first phone safety contact will start 13 weeks after V6 as applicable. Patients will be contacted at least 1 week before each visit to be reminded of procedures and study drug administration.

In order to maintain the blind, ALP, gamma-glutamyltransferase (GGT) and 5' nucleotidase obtained from blood samples from V2 to LVDB will be kept in blinded condition during the Double-blind period for each patient. Any blinded party will not be informed of these values until the database freezing and the Sponsor authorization to unblind the data at the end of the DB period.

If a patient prematurely discontinues the study drug during the DB period, end of treatment (EOT) DB Visit should be performed between 16 and 30 days after last drug intake. The follow-up of the patients who discontinue the drug early will continue until W104, or until the last completed V6 whichever occurs first, and will complete all visit procedures except liver biopsy and PK (if already done prior to study drug discontinuation). The study procedures will be assessed as per Table 1 and Table 2. At the scheduled visit following early discontinuation/termination of the study drug, the date and the reason of the study drug discontinuation will be collected. In case the EOT visit occurs within the time window of the next scheduled visit, EOT visit will be performed at the scheduled visit.

Only patients who completed the DB period (excluding patients who permanently stopped the study drug

during the DB period) will be included into the open label LTE.

2.5 Long-Term extension

The second study period is an open-label LTE. During this LTE period, all patients will receive 80 mg of Elafibranor for up to 5 years, so that the total duration of study participation for a patient is up to 6 years. Apart from the first two visits in this LTE period, which will occur over a 13-week interval starting after V6, V8 or the LVDB, patients will return to the site every 26 weeks (+/-14 days) for procedures listed on Table 1 and Table 2, and safety contacts will be performed by phone call between on-site visits, every alternating 26 weeks (+/-14 days), to collect information related to AEs, concomitant medications, pregnancy, and study drug compliance. The Investigator may subsequently decide to perform an unscheduled visit.

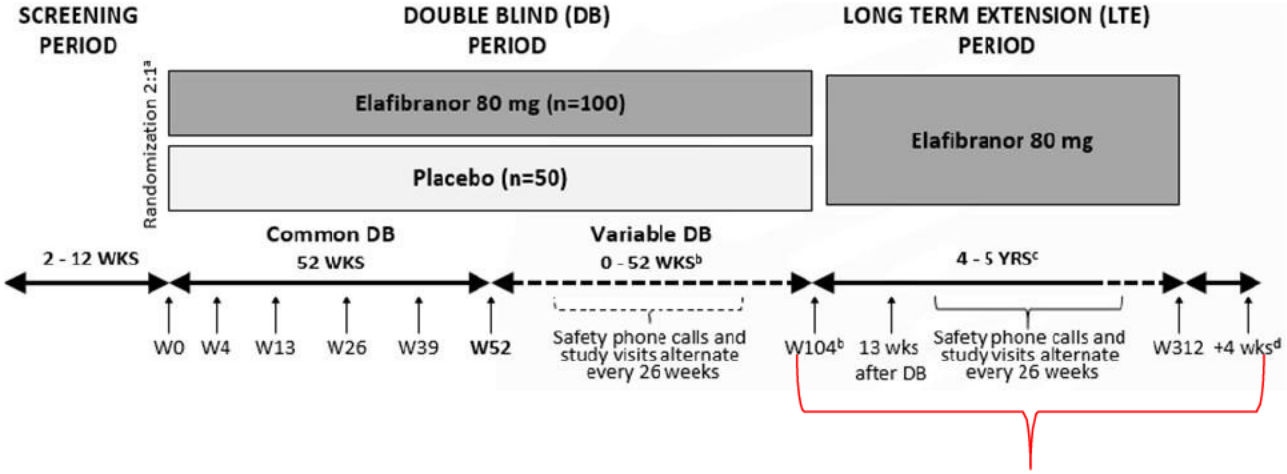
The statistical analyses of the LTE phase are not within the scope of this SAP and will be described in a separate SAP. The LTE data will be included in the SDTM domains but will not be part of the ADAM datasets nor the statistical outputs of the double-blind period.

During the open label LTE, an EOT LTE visit will be performed between 16 to 30 days after the last drug intake for any patient who discontinue study drug. No further follow-up will be required. The study procedures will be assessed as per Table 1 and Table 2.

2.6 Clinical event committee (CEC)

The CEC will conduct adjudication of the clinical outcomes and DILI events. The CEC assessment and adjudication will occur in a blinded (during DB period) and consistent and unbiased manner throughout the course of the study to determine whether the event meets the protocol specified criteria. The CEC will be comprised of 3 hepatologists, and all of them will be independent of the conduct of the study.

Figure 1: Study Design Overview



Footnotes:

The statistical analyses of the Long-Term Extension period are not covered in this SAP.

a. If receiving UDCA at randomization, continue throughout studyparticipation

b. The Variable DB duration is an additional 52 weeks after end of Common DB (W104) or until the last completed W52 / V6 (W52), whichever occurs first

c. The LTE duration is 5 years after end of the DB period or until the patient's total treatment duration is 6 years, whichever occurs first

d. Safety follow-up 4 weeks after last dose of study drug

2.7 Optional Visits

2.7.1 Retesting and/or additional screening visits

- If creatine phosphokinase (CPK) value is > 2xULN at SV1, it can be repeated within 1 to 2 weeks, but prior to V1.
- If HCV Ab test is positive at SV1, a patient's HCV infection status needs to be confirmed by HCV RNA testing prior to V1.

Any other retest deemed necessary by the Investigator should be discussed with the Study Medical Monitors.

2.7.2 Unscheduled visits

An unscheduled visit is defined as any visit to the study unit outside of the protocol-evaluation timepoints where the patient is seen by study unit personnel, e.g., when follow-up assessments are required for safety reasons or when repeat measurements are required out of the screening period (either to confirm a measurement or in case of errors, measuring device failure, etc.).

2.8 Exploratory/Ancillary Sub-study

No Exploratory/Ancillary Sub-study is planned.

2.9 Randomization Methodology

Patients who satisfy all eligibility criteria will be randomized in a 2:1 ratio to one of the following groups:

- Elafibranor 80 mg
- Placebo

A central randomization system will be used (interactive voice/web response system [IxRS]).

The randomization will be stratified at V1 (randomization/baseline visit = V1) on two factors, collected during the screening period:

- ALP > 3x ULN or Total bilirubin > ULN: Yes/No, and
- PBC Worst Itch NRS averaged - over the 14 days preceding randomization - ≥ 4: Yes/No.

To ensure inclusion of a relevant ratio of patients with moderate disease or increased risk of progression:

- approximately 10% of the total randomized patients will present with a TB > ULN or Albumin (ALB) < lower limit of normal (LLN) (definition of moderately advanced patients per Rotterdam criteria))

and

- approximately 20% of the total randomized patients will present with a TB > 0.6 x ULN.

2.10 Stopping Rules and Unblinding

An independent Data Safety Monitoring Board (DSMB) composed of relevant experts (an endocrinologist, cardiologist, hepatologist and nephrologist) experienced in the management of patients with PBC and an independent biostatistician will oversee the study conduct. The DSMB will be established in order to review on a regular basis during the study (at least every six months) the progress of the study and the safety of the treatment in an unblinded manner, to protect patient welfare and preserve study integrity. A first DSMB will be planned within the first 6 months after First Patient randomized First Visit.

The Chief Medical Officer (CMO) will be responsible for promptly reviewing the DSMB recommendations and determining whether expedited reporting of any safety issues, amendments to the protocol, or changes in study conduct are required.

The study blind may be broken for an individual patient only in the case of an emergency when knowledge of the study drug is essential for the clinical management of the patient. The randomization number allocated to the patient will allow any unblinded study personnel to identify the treatment group to which the patient is randomized.

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

2.11 Study Procedures

The schedule of assessments, as outlined in the study protocol, is provided in Table 1 and in Table 2.

Table 1: Schedule of Assessments

Study Period	Screening			Double-Blind (DB)									If applicable	Safety contact in variable DB & LTE[l]	Long-term Extension (LTE)		
				Common DB						Variable DB							
Visit number	SV1	SV2	SV3	V1	V2	V3	V4	V5	V6	V7	V8	LVDB[j]	EOT DB[k]		LT1	LT2 to LTn	EOT-LTE[k]
Weeks	-12 to -2			0	4	13	26	39	52	78	104	At max. 13 weeks after last V5 for the last patient	Safety Follow-up: 16 to 30 days after last drug intake	DB period: -Week 65 -Week 91 LTE: - 13 weeks after LT2 then every 26 weeks	13 weeks after last DB visit	LT2 13 weeks after LT1 then every 26 weeks	Safety Follow-up: 16 to 30 days after last study drug intake
Days				1	29	92	183	274	365	547	729	NA	NA	Week 65: 456 days Week 91: 638 days LTE: 91 d after	V6 or V8 or LVDB + 91 d	LT2: 91 days after LT1 then every 182	NA

CCI [redacted]

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

													LT2 then every 182 d		days up to 2185	
Tolerances (Days)					+/- 7	+/- 14	+/- 14	+/- 14	+/- 14	+/- 14	+/- 14	NA	NA	+/- 14	+/- 14	+/- 14	NA
STUDY PROCEDURES [a]																	
Obtain Informed Consent	X																
Medical and Disease History	X																
Inclusion/Exclusion Criteria	X			X													
Physical Examination (Height at SV1 only)	X			X	X	X	X	X	X	X	X	X	X		X	X	X
Vital Signs and Weight	X			X	X	X	X	X	X	X	X	X	X		X	X	X
12-lead Electrocardiogram	X			X			X		X		X					X	
PRO questionnaires[b]																	
PBC Worst Itch NRS	X[f] ►																

CCI [redacted]

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	Document Date:	10-May-2023

PBC Worst Itch NRS-Past Week								X	X	X	X			X	
PGIC			X	X	X	X	X	X	X	X	X			X	
PGIS	X[g]	X	X	X	X	X	X	X	X	X	X			X	
5-D Itch		X	X	X	X	X	X	X	X	X	X			X	
PROMIS Fatigue Short Form 7a		X	X	X	X	X	X	X	X	X	X			X	
ESS		X	X	X	X	X	X	X	X	X	X			X	
PBC-40		X	X	X	X	X	X	X	X	X	X			X	
EQ-5D-5L		X	X	X	X	X	X	X	X	X	X			X	
Transient Elastography (TE) (Fibroscan)		X			X		X	X	X	X				X	
Liver Biopsy (optional)	X[h]						X								
Ultrasound exam (liver & bladder) [c]	X						X		X					X	
Hip and lumbar DEXA scanning [d]	X						X							X	

CCI [redacted]

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	Document Date:	10-May-2023

PK			X[i]											
Adverse Events (AEs)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Clinical Outcomes		X	X	X	X	X	X	X	X	X	X	X	X	X	X
Concomitant Medications	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Randomization		X													
Treatment Assignment [e]		X					X								
Dispense Study Drug		X		X	X	X	X	X	X	X				X[m]	
Study Drug Accountability/Compliance			X	X	X	X	X	X	X	X	X			X	X
EOS registration		To be completed in study system(s) as applicable													

Footnotes:

a. Procedures/assessments should be conducted in the following order during study visits: PROs (when completed at the study center), investigator assessments, safety and laboratory assessments, administration of study drug

b. Refer to Table 3 (Schedule of PRO Questionnaires) for details

c. At baseline, US exam can be performed before randomization after the patient has been otherwise confirmed as eligible and up to randomization visit. Ultrasound exam to be performed every year and for any visit post baseline US exam can be performed with a tolerance of +/ 7 days around the planned visit date.

d. Hip and lumbar DEXA scanning to be performed in all patients where the exam is accessible to sites at baseline, at V6, and then 2 years later. At baseline, the exam can be performed before randomization after the patient has been otherwise confirmed as eligible and up to randomization visit. For exam post baseline, DEXA exam can be performed with a tolerance of +/ 7 days around the planned visitdate.

CCI [redacted]

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

e. The switch to elafibranor treatment will happen either at V8 or at LVDB (either V7 or V8, depending on when the last patient in the study completes his/her V5).
f. During the screening period and DB period up to week 52 (V6), the PBC Worst Itch NRS score will be collected every evening via an eDiary. The mean score of the 14 days prior to randomization (V1) will be used for stratification. Patients must have at least 4 available values for PBC Worst Itch NRS during each of the 7-day intervals in the 14-day intervals prior to V1, for a total of at least 8 values in the 14 days prior to V1.
g. Patient Global Impression of Severity (PGIS) will be collected at SV1 only during the screening period.
h. For patients who have consented to have liver biopsy samples collected, Liver biopsy at inclusion to be done at any SV preferably in patients already confirmed as eligible and if at all possible 2 to 4 weeks prior to randomization. LB at week 52 can be performed with a tolerance of +/- 2 weeks around the planned visit V6.
i. PK assessment will include the following timepoints: pre-dose, 0.5h, 1.5h, between 2 and 3h, 4h, and 6h.
j. Until the last patient in the study completes his/her V5, patients between V6 and V8, will complete LVDB according to Table 1 General Assessment Schedule. After the last patient in the study completes his/her V5, patients between V6 and V8, will complete LVDB at the next scheduled visit (either V7 or V8). LVDB will replace V7 or V8. The same procedures as for V8 will be performed for LVDB (except US exam). For patients who have not yet had V6 at the time the last patient completes his/her V5, LVDB will be scheduled at the latest 13 weeks after the date on which V5 was completed for the last patient. For those patients, LVDB will coincide with V6, and patients will complete V6 and associated procedures to facilitate transition to the open-label elafibranor treatment phase in the long-term extension study, in a timely manner.
k. Safety contact by phone call every alternating 26 weeks starting 13 weeks after V6 in the DB period and starting after LT2 during LTE to check AEs and concomitant medications
l. If premature study drug discontinuation during DB period, EOT DB Visit should be performed between 16 and 30 days after last drug intake, and patients will continue in the study until V8, or until the last completed V6 whichever occurs first. In case the EOT DB visit occurs within the time window of the next scheduled visit, EOT DB visit replaces the scheduled visit. If premature study drug discontinuation occurs during LTE period, an EOT LTE visit will be performed between 16 and 30 days after last drug intake
m. Drug dispensation will be done up to LT10. There will be no drug dispensation at LT11

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

Table 2: Schedule of Biological Assessments

Study Period	Screening			Double-Blind (DB)									If applicable	Long-term Extension (LTE)		
				Common DB						Variable DB						
Visit number	SV1	SV2	SV3	V1	V2	V3	V4	V5	V6	V7	V8	LVDB	EOT DB[i]	LT1	LT2 to LTn	EOT-LTE[b]
Weeks	-12 to -2			0	4	13	26	39	52	78	104	At max. 13 weeks after last V5 for the last patient	Safety Follow-up: 16 to 30 days after last drug intake	13 weeks after last DB visit	LT2 13 weeks after LT1 then every 26 weeks	Safety Follow-up: 16 to 30 days after last study drug intake
Days				1	29	92	183	274	365	547	729	NA	NA	V6 or V8 or LVDB + 91 d	LT2: 91 days after LT1 then every 182 days up to 2185	NA
Tolerances (Days)					+/- 7	+/- 14	+/- 14	+/- 14	+/- 14	+/- 14	+/- 14	NA	NA	+/- 14	+/- 14	NA
Serum Hematology *Hemoglobin, hematocrit, WBC with differential, platelet count, prothrombin time (Pt), INR, reticulocytes/RBC*	X			X	X	X	X	X	X	X	X	X	X	X	X	X
Screening Serum Chemistry [a]	X															

CCI [redacted]

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	Document Date:	10-May-2023

ALP, ALT, AST, GGT, CPK, total and conjugated bilirubin, albumin, creatinine, sodium, AFP, eGFR (MDRD formula & CKD-EPI formula), MELD-Na														
Screening serum hCG pregnancy test [b]	X													
Serum Chemistry *Sodium, potassium, chloride, calcium, albumin, BUN, creatinine, TB, conjugated bilirubin, AST, ALT, ALP, GGT, 5' NT, total proteins, lipase, amylase, TC, LDL-C, HDL-C, VLDL-C, TG, CPK, FPG, eGFR, MELD-Na*		X	X	X	X	X	X	X	X	X	X	X	X	X
ALP fractionated		X					X							
Serology *HIV Ab I/II, Anti- HAV IgM, HBs, HCV*	X													
Serum Bile Acids and Biomarkers of Bile Acid Synthesis *Bile acids (cholic acid (CA), glycocholic acid (GCA), taurocholic acid (TCA), chenodeoxycholic acid (CDCA), glycochenodeoxycholic acid (GCDCA), taurochenodeoxycholic acid (TCDCA), deoxycholic acid (DCA), glycodeoxycholic acid (GDCA), taurodeoxycholic acid (TDCA), lithocholic acid (LCA), glycolithocholic acid (GLCA), taurolithocholic acid (TLCA)), C4, FGF-19*		X			X		X	X	X	X			X	

CCI [redacted]

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	Document Date:	10-May-2023

Biomarkers of Hepatic Fibrosis and/or Inflammation *HsCRP, fibrinogen, haptoglobin, TNF-α, IL-6, ELF (HA, PIINP, TIMP-1), PAI-1, TGF-β, CK-18 (M65 and M30), Pro-C3*		X			X		X	X	X	X			X	
Additional Safety Markers *Cystatin C, urine albumin to creatinine ratio (urine ACR), AFP* [c]		X			X		X	X	X	X	X		X	X
Immunoglobulins *IgG, IgM*		X			X		X	X	X	X			X	
Serum Bone markers [d] *CTX, P1NP*		X		X	X		X							
Urinalysis (dipstick)[e] *Specific gravity, pH, protein, glucose, ketones, bilirubin, urobilinogen, blood, nitrite, leukocytes*	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Urine-based β-human chorionic gonadotropin (hCG) Pregnancy Test [f]		X	X	X	X	X	X	X	X	X		X	X	
urinary myoglobin, serum IgG and SMA [g]		X	X	X	X	X	X	X	X	X	X	X	X	X
Biobank (optional) [h]		X			X		X	X	X	X			X	

Footnotes:

a. Repeated measured for AST, ALT, ALP and TB to be collected (See section 2.2).

b. Serum pregnancy test must be performed at screening in all females of childbearing potential and may be repeated within one month prior to randomization in case the screening period lasts more than 4 weeks.

c. AFP to be valuated at V1, V6, and then everyyear, as well as at LVDB, if applicable.

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

d. Serum bone markers to be assessed at baseline, week 13, week 26, and week 52
e. Microscopic evaluation will be performed if dipstick urinalysis indicated presence of any significant abnormality.
f. In all females of childbearing potential, urine-based β-hCG pregnancy tests at any site visits from V1. In between site visits, a home pregnancy test is to be performed every 4 weeks, starting after V1. If the urine-based test is positive, a confirmatory serum pregnancy test must be performed at site.
g. Assessment of presence of myoglobin in urine to be done locally only in case of clinically significant CPK elevation; assessment of IgG and SMA to be done locally in case of suspicion of AIH
h. Additional blood samples will be collected from patients, who have given their consent, to be used to discover or validate biomarkers in PBC and related diseases.
i. If premature study drug discontinuation during double-blind period, EOT DB Visit should be performed between 16 and 30 days after last drug intake, and patients will continue in the study until V8. In case the EOT DB visit occurs within the time window of the next scheduled visit, EOT DB visit replaces the scheduled visit. If premature study drug discontinuation occurs during LTE period, an EOT LTE visit will be performed between 16 and 30 days after last drug intake.

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

Table 3: Schedule of Patient Reported Outcomes Questionnaires

Platform	Assessment	Frequency and Duration of Assessment	Time of assessment
eDiary	PBC Worst Itch NRS	Once daily during screening[a] and the Common DB period (up to V6)	Evening
eTablet	PBC Worst Itch NRS Past Week	Throughout the Variable DB and LTE periods	During study visit[b]
eTablet	PGIC	Throughout the DB (starting at V2) and LTE periods	During study visit[b]
eTablet	PGIS	Throughout the screening, DB, and LTE periods	During study visit[b]
eTablet	5-D Itch	Throughout the DB and LTE periods	During study visit[b]
eTablet	PROMIS Fatigue Short Form 7a	Throughout the DB and LTE periods	During study visit[b]
eTablet	ESS	Throughout the DB and LTE periods	During study visit[b]
eTablet	PBC-40	Throughout the DB and LTE periods	During study visit[b]
eTablet	EQ-5D-5L	Throughout the DB and LTE periods	During study visit[b]

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

Footnotes:

a. The mean PBC Worst Itch NRS score of the 14 days prior to randomization (V1) will be used for stratification. Patients must have at least 4 available values for PBC Worst Itch NRS during each of the 7 day intervals in the 14 days prior to V1, for a total of at least 8 values in the 14 days prior to V1.
b. Administered on site.

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	Document Date:	10-May-2023

2.12 Efficacy, Pharmacokinetic, and Safety Variables

2.12.1 Efficacy Variables

The efficacy of Elafibranor (80 mg/day) compared to placebo on cholestasis will be evaluated considering the response to treatment in terms of:

Primary endpoint:
ALP < 1.67x ULN **and** TB ≤ ULN **and** ALP decrease from baseline ≥ 15% at week 52 (binary).

Furthermore, the following variables will be considered for the evaluation of efficacy:

Three key secondary endpoints:

- Response to treatment based on ALP normalization at week 52 (binary).
- Change in pruritus from baseline through week 52 on PBC Worst Itch NRS in patients with baseline PBC Worst Itch NRS score ≥4 (continuous).
- Change in pruritus from baseline through week 24 on PBC Worst Itch NRS in patients with baseline PBC Worst Itch NRS score ≥4 (continuous).

Other Secondary Endpoints:

- Change from baseline in ALP at 4, 13, 26, 39 and 52 weeks (continuous)
- ALP response definedas 10%, 20% and 40% ALP reduction from baseline at week 52 (binary)
- Response to treatment at week 52 according to (all binary):
  a) ALP < 1.5x ULN, ALP decrease from baseline ≥ 40% and TB ≤ ULN
  b) ALP < 3x ULN, AST < 2x ULN and TB < 1mg/dL (Paris I)
  c) ALP ≤ 1.5x ULN, AST ≤ 1.5x ULN and TB ≤ ULN (Paris II)
  d) TB response rate of 15% change from baseline
  e) Normalization of abnormal TB (TB ≤ ULN) and/or ALB (ALB ≥ LLN) (Rotterdam)
  f) TB ≤ 0.6x ULN
  g) ALP ≤ 1.67x ULN and TB ≤ 1mg/dL (Momah/Lindor criterion, 2011)
  h) No worsening of TB defined as level of TB < ULN or no increase from baseline of more than 0.1x ULN
  i) Complete biochemical response defined as normal ALP, TB, AST, ALT, albumin, and INR
- PBC risk scores at week 52: UK PBC score (Carbone_et_al_2016) and GLOBE score (a risk score to predict transplantation-free survival at 5, 10 and 15 years. (Lammers_et_al_2015))

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	Document Date:	10-May-2023

- Response to treatment based on the normalization of bilirubin (TB ≤ ULN) at week 52 (binary)
- Response to treatment based on the normalization of albumin (ALB ≥ LLN) at week 52 (binary)
- Change from baseline to week 52 in hepatobiliary injury and liver function as measured by AST, ALT, GGT, 5' NT, total and conjugated bilirubin, albumin, INR and ALP fractionated (hepatic) (continuous)
- Change from baseline to week 52 in biomarkers of inflammation as measured by hsCRP, fibrinogen, haptoglobin and TNF-α (continuous)
- Change from baseline to week 52 in immune response as measured by IgG and IgM (continuous)
- Change from baseline to week 52 in biomarkers, and non-invasive measures of hepatic fibrosis as measured by ELF test (HA, PIINP, TIMP-1), PAI-1, TGF-β, CK-18 (M65 and M30), Pro-C3 and liver stiffness measured by Transient Elastography (continuous)
- Change from baseline to week 52 in lipid parameters as measured by TC, LDL-C, HDL-C, calculated VLDL-C and TG (continuous)
- Change from baseline to week 52 in Fasting Plasma Glucose (continuous)
- Change from baseline to week 52 in bile acids and biomarkers of bile acid synthesis as measured by bile acids, C4 and FGF-19 (continuous)
- Proportion of responders in PBC Worst Itch NRS according to clinically meaningful change; at least 30% reduction; and one point, two points or three points decrease in score from baseline through week 52 and through week 24 in patients with a baseline NRS score ≥ 4 (binary)
- Proportion of participants achieving sustained improvement in PBC Worst Itch NRS as defined as having responses for all the last three 4-week periods (periods 11, 12 and 13) according to clinically meaningful change; at least 30% reduction; and one point, two points or three points decrease in score from baseline in patients with a baseline NRS score ≥ 4 (binary)
- Proportion of patients with no worsening of pruritus from baseline through week 52 and through week 24 as measured by the PBC Worst Itch NRS (binary)
- Change from baseline to week 52 in 5D-Itch (continuous)
- Change from baseline to week 52 in PROMIS Fatigue Short Form 7a(continuous)
- Change from baseline to week 52 in ESS (continuous)
- Change from baseline to week 52 in PBC-40 (continuous)
- Change from baseline to week 52 in health utility as measured by the EQ-5D-5L(continuous)
- Change from baseline to week 52 in serum markers of bone turnover and in bone mineral density (hip and lumbar) assessed by DEXA scanning (continuous)
- Onset of clinical outcomes, described as a composite endpoint (binary) composedof:

a) MELD-Na > 14 for patients with baseline MELD-Na < 12

b) Liver transplant

c) Uncontrolled ascites requiring treatment

d) Hospitalization for new onset or recurrence of any of the following:

i) Variceal bleed

ii) Hepatic encephalopathy defined as West-Haven/Conn score of 2 or more

iii) Spontaneous bacterial peritonitis

e) Death

2.12.2 Pharmacokinetic Variables

The statistical considerations applicable to the popPK modeling, including the definition of the PK population and the description of the statistical analyses, will be fully described in a dedicated Analysis Plan (popPK SAP).

2.12.3 Safety Variables

The safety and tolerability profiles of Elafibranor (80mg/day) will be assessed from the recordings of:

a) Serious adverse events (SAEs), adverse events (AEs), and adverse events of special interest (AESIs),

b) Physical examination findings, vital signs (including weight), medical history, 12-lead Electrocardiogram (ECG), ultrasound examination of liver and bladder/urinary tract,

c) Serum chemistry (Sodium, potassium, chloride, calcium, ALB, BUN, creatinine, fasting plasma glucose, total proteins, lipase, amylase, TC, LDL-C, HDL-C, VLDL-C, TG, CPK, eGFR, MELD-Na) and hematology (Hemoglobin, hematocrit, WBC with differential, platelet count, PT),

d) Hepatobiliary injury and liver function markers AST, ALT, GGT, 5' NT, total and conjugated bilirubin, ALB, International normalized ratio (INR) and ALP fractionated (hepatic),

e) Renal biomarkers (including urinalysis: Specific gravity, pH, protein, glucose, ketones, bilirubin, urobilinogen, blood, nitrite, leukocytes), and

f) Other biochemical safety markers (Cystatin C, urine ACR, urinary myoglobin) and biomarkers of inflammation (HsCRP, fibrinogen, haptoglobin, TNF-α, IL-6, ELF test (HA, PIINP, TIMP-1), PAI-1, TGF-β, CK-18 (M65 and M30), Pro-C3).

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

The planned statistical analyses related to the open label LTE period are not within the scope of this SAP and will be described in a separate SAP.

CCI

[redacted]

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	Document Date:	10-May-2023

3. SUBJECT POPULATIONS

3.1 Population Definitions

The following patient populations will be used for the analysis and the presentation of the data:

Screened Analysis Set: All patients who sign informed consent. This population will be used to summarize disposition.

Intent-to-treat (ITT) Analysis Set: All randomized patients. Patients will be analyzed according to randomized treatment.

Pruritus ITT Analysis Set: All patients from the ITT analysis set with baseline PBC Worst Itch NRS score ≥4.

Per Protocol (PP) Analysis Set: All patients from the ITT population without any major protocol deviation or event affecting the primary efficacy endpoint (see section 3.2).

Pruritus PP Analysis Set: All patients from the Pruritus ITT analysis set without any major protocol deviation or event affecting the primary efficacy endpoint and/or the second and third key secondary endpoint.

Safety Analysis Set: All patients who were administered at least one dose of DB study drug irrespective of the treatment received. Patients will be analyzed according to the treatment received. Patients who received any amount of active treatment, even by mistake and for one intake, will be assigned to the active treatment group.

Pharmacokinetics Analysis Set (PKS): All patients who were administered at least one dose of Elafibranor and have at least one post-dose PK sample. Patients of the PKS must have data for time of dosing, time of sampling and amount of product administered. Whereas all patients are sampled in order to maintain the blind, the PKS will only include patients under Elafibranor.

CCI [redacted]

The primary and first key secondary efficacy analyses will be performed primarily on the ITT analysis set, while only the main analyses of the primary and first key secondary endpoints will be replicated on the PP analysis set.
The second and third key secondary analyses will be performed primarily on the pruritus ITT analysis set, while only the main analyses of the second and third key secondary endpoints will be replicated on the pruritus PP analysis set.

Each efficacy endpoint will be evaluated up to week 52. For patients who completed additional visits during the DB period, descriptive statistics will be presented up to the end of the DB period.

The Safety Set will be typically the primary analysis set for the analysis of safety endpoints.

CCI [redacted]

CCI [redacted]

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

CCI [redacted]

If both ITT and safety analysis sets are the same, meaning that all randomized patients have taken at least one study treatment dose and none are reallocated to a different treatment group compared to randomization, the replicated analysis planned on safety analysis set for demographics, screening and baseline characteristics will not be done.

3.2 Protocol Deviations

All protocol deviations will be entered in a Risk and Issue Management System (RIM) and managed according to the Protocol Deviation Plan (PDP). They will be classified as non-important or important based on their effect on the rights, safety, or well-being of the patients and/or the quality and integrity of the data, and the final rating of all deviations will be confirmed on a case-by-case basis and prior to database lock (DBL). Important deviations do not necessarily lead to the exclusion from the PP population. Some examples of protocol deviations are given below:

- Protocol deviations related to Inclusion/Exclusion criteria (ineligible patient randomized to the study)
- Randomization/drug dispensation errors (incorrect kit dispensed to apatient)
- Treatment compliance less than 80% or greater than 120% from baseline to Week 52
- Any other protocol deviations that may affect patient safety, data integrity, or future conduct of the study.

The sponsor, or designee, will be responsible for producing the final protocol deviations file (formatted as a Microsoft Excel file), in collaboration with Cytel and the data monitoring group as applicable, and for determining in a blinded way what kind of deviations will require the corresponding data to be excluded from the PP population according to the Per Protocol Set Plan.

This file will include a description of the protocol deviation, and clearly identify whether this deviation warrants exclusion from the PP/pruritus PP populations or not. This file will be finalized as well as the list of events that exclude from PP/pruritus PP (according to the Per protocol set definition plan) prior to hard database lock at the end of the DB period and completed prior to the DB hard database lock for the primary endpoint analysis.

No update will be allowed after the DB database lock on the data collected up to the Last Patient Last Visit performed during DB period (V6/V8/LVDB, Safety Follow-up 16-30 days after last dose of study drug).

All protocol deviations will be presented in data listings. All important deviations and deviations that lead to the exclusion from the PP population will be summarized in a table.

Note: Major/Minor and Important/Non-important have been used interchangeably in the study capturing the exact same notion.

4. STATISTICAL METHODS

4.1 Sample Size Justification

CCI [redacted]

One hundred and fifty patients (100 Elafibranor vs 50 placebo) allow to achieve at least 90% power to demonstrate a statistically significant between group difference of 35% (47% in Elafibranor group vs 12% in placebo group) in the response rate at week 52 of the primary efficacy endpoint with a two-sided alpha of 0.05 and using an exact Fisher test.

Assuming 1/50 (2.0%) patients in the placebo arm with ALP normalization at week 52 (first key secondary endpoint), 150 patients (100 Elafibranor vs 50 placebo) provide at least 80% power to detect a statistically significant between group difference of 20.0% at a two-sided 0.05 alpha level.

Assuming a pooled standard deviation of 2.3 points, 60 patients (40 Elafibranor vs 20 placebo) with baseline PBC Worst Itch NRS score ≥4 provide approximately 80% power to detect a statistically significant between group difference of 1.8 points in mean change from baseline in PBC worst itch NRS score (second key secondary endpoint) at a two-sided 0.05 alpha level. It is assumed that the same assumptions would apply to the two key secondary endpoints for pruritus (through week 52 and through week 24).

4.2 General Statistical Methods and Data Handling

4.2.1 General Methods

All output will be incorporated into Rich Text Format (RTF) files, sorted and labeled according to the International Conference on Harmonization (ICH) recommendations, and formatted to the appropriate page size(s).

Statistical methods will focus on summarizing the data collected using appropriate graphical and tabular presentations and on generation of inferential summaries.

Tabulations will be produced for appropriate demographic, baseline, efficacy and safety parameters.

For measurements of continuous endpoints, summary statistics of absolute values and absolute/relative changes from baseline, where appropriate, will be summarized and will include n, mean, standard deviation, median, 25th percentile - 75th percentile (Q1-Q3), and minimum and maximum values. The number of missing values will be displayed in parenthesis next to 'n'. Mean, standard deviation, median, Q1 and Q3 will be presented with one more decimal place compared to the raw data, and minimum and maximum will be presented with the same number of decimal places as the raw data.

CCI [redacted]

For categorical variables, summary tabulations of the number and percentage within each category of the parameter will be presented (as well as the number for missing data). The total number of non- missing patients will be used as the denominator. Percentages will be rounded to one decimal place. Therefore, there may be cases where for instance the total of the percentages does not exactly equal 100%. Where applicable confidence intervals (95% CI for primary, key secondary endpoints and for all other secondary endpoints, 2-sided) will be presented, and they will be shown with 1 decimal place for percentages, and with 2 decimal places for continuous measures.

Formal statistical hypothesis testing will be performed on the primary and key secondary efficacy endpoints, with all tests conducted at the 2-sided, 0.05 level of significance. The fixed-sequence testing approach will be used to control the overall type I error rate at a two-sided 0.05 level. Summary statistics will be presented, as well as CIs on selected parameters, as described in the sections below.

The fixed-sequence testing approach implies that if the primary endpoint is statistically significant at a two-sided 0.05 level, the first key secondary endpoint (ALP normalization) will be tested at the same level. If the first key secondary endpoint is statistically significant at a two-sided 0.05 level, the second key secondary endpoint (change in pruritus through week 52) will be tested at the same level. If the second key secondary endpoint is statistically significant at a two-sided 0.05 level, the third key secondary endpoint (change in pruritus through week 24) will be tested at the same level. If the primary endpoint is NOT statistically significant, none of the key secondary endpoints will be tested. If the primary endpoint is statistically significant and the first key secondary endpoint is NOT statistically significant, the second and third key secondary endpoints will not be tested. Finally, if the second key secondary endpoint is NOT statistically significant, the third key secondary endpoint will not be tested.

Statistical testing for other secondary endpoints will be of exploratory nature and will be performed at the two-sided 0.05 level of significance.

All summarized data will be listed in addition to the analyses and summaries described below.

Analyses described in this SAP will be dependent on the availability of the data collected. Additional exploratory analyses may be performed as well as those described below, if suggested by inspection of the final database.

4.2.2 Data Conventions

In the context of this SAP, we will consider 4 periods (max. 12 weeks screening, max. 104 weeks DB, max. 260 weeks LTE, 4 weeks follow-up; only the statistical analyses of the first two periods are detailed in this SAP):

1) Screening period, defined as the period prior to Study Visit 1 (before first study drug dispensation), which could last from 2 to 12 weeks.

2) Double-blind treatment period of the study, defined as the time between Study Visit 1 (first study drug dispensation) and the last completed week 52 (V6) or a maximum of 104 weeks DB period, whichever happens first.

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	Document Date:	10-May-2023

The double-blind period includes the screening period and the double-blind treatment period, and its start date will be the first screening visit date.

The double-blind period end date will be derived according to the patient status at the time of database lock:

– if the patient discontinued the study treatment or the study, it will be the last available date from all domains or the death date.
– For other patients for whom rollover to long-term extension period is assumed:
  - if patient doesn't have first intake date of long-term extension period then it will be the last double-blind visit date;
  - if patient has first intake date of long-term extension period then it will be the maximum date of the first intake date of long-term extension period -1 and the last double-blind visit date.

Note that the last double-blind visit date is the maximum of the last visit date including visit 6, visit 7, visit 8, and last visit double-blind.

3) Long-term extension period of the study, defined as the time between the Week 52/LVDB/week 104 visit and the Week 312 visit (max. 260 weeks for the LTEperiod)

4) Follow-up period, defined as the period beyond the Week 312 visit or beyond the study treatment discontinuation happening during LTE period (follow-up period allowing 4 weeks for safety follow-up).

The study period will end after approximately 6 years from SV1, but this SAP only provides the details for the DB (12 weeks of screening and up to 104 weeks of DB period) analyses.

The week 0 Visit 1 (V1) is defined as the first treatment visit, where the patient is randomized and given the study Treatment carton (kit) for the first time.

Treatment day 1 is defined as the date of first study drug intake of any of the following study drugs: Placebo or Elafibranor 80mg.

The EOT DB visit (in case of study drug discontinuation during the DB period), is defined as the safety visit date recorded during Double-blind period.

Either the V6, LVDB, V7 or V8 can be defined as the latest visit date recorded during Double-blind period.

The Week 312 visit (LT9/LT10/LT11) is defined as the latest date from any panel of Long-term extension period.

The EOT LTE visit date is defined as the last recorded visit date from any panel of the Follow-up period.

The following conversion factors will be used to convert days to months or years, or Fahrenheit to Celsius, where applicable:

- 1 month = 30.4375 days

- 1 year = 365.25 days
- 1 week = 7 days
- °C = (°F - 32)/1.8.

Additional data handling rules are as follows:

- Age (years) = (date of informed consent – date of birth + 1) / 365.25, rounded to the lowest whole number.
- Weight values recorded in pounds will be converted to kilograms using the following formula: kilograms = pounds/2.2046.
- Height values recorded in inches will be converted to centimeters using the following formula: centimeters = inches*2.54.
- Duration on study (weeks) = (Last visit date – randomization date + 1) / 7.
- Duration on treatment (weeks) = (Last study drug intake date – first study drug intake date +1 – days of treatment interruption) / 7.
- Duration of exposure (weeks) = (Last study drug intake date – first study drug intake date + 1) / 7.
- (Absolute) Change from baseline = Value at the time point – Baselinevalue.
- Relative change from baseline = [(Value at the time point – Baseline value) / Baseline value] x 100.

As a general rule of thumb, each laboratory value below/above the limit of quantification will be imputed numerically by the limit of quantification.

4.2.3 Computing Environment

All descriptive statistical analyses will be performed using SAS statistical software version 9.4. Medical history, Adverse Events and Procedures will be coded using version 26.0 of the Medical Dictionary for Regulatory Activities (MedDRA). Concomitant medications will be coded in the same way using World Health Organization (WHO) Drug dictionary version B3 WHO Drug global March 2023.

4.2.4 Methods of Pooling Data

N/A.

4.2.5 Adjustments for Covariates

The randomization will be stratified across two factors: (ALP > 3x ULN or total bilirubin > ULN (Yes/No) and PBC Worst Itch NRS score averaged - over the 14 days preceding baseline - value ≥ 4) (Yes/No) at baseline

(V1) (see section 2.9).

All the categorical endpoints will be analyzed using the exact Cochran-Mantel-Haenszel (CMH) test stratified by the randomization factors.

The sensitivity analysis to the statistical model (see section 4.6.2.2) performed on the primary endpoint and the first key secondary endpoint will be analyzed using an exact logistic regression model stratified by randomization strata with treatment group as factor, on both ITT and PP populations.

The continuous endpoints will be compared between treatment groups using the Mixed effect Model for Repeated Measurement (MMRM), including the treatment indicator, visit and stratification factors as fixed factors, interaction of treatment and visit and the respective baseline value as covariate.

4.2.6 Multiple Comparisons/Multiplicity

Multiplicity will be dealt with through the Fixed-sequence Testing Procedure that implies a pre- specified sequence of hypotheses testing. In this study the first key secondary endpoint will only be tested if the primary endpoint is statistically significant, the second key secondary endpoint (change in pruritus through week 52) will only be tested if the first key secondary endpoint is statistically significant and the third key secondary endpoint (change in pruritus through week 24) will only be tested if the second key secondary endpoint is statistically significant (see section 4.2.1).

4.2.7 Subpopulations

Exploratory analyses of the primary endpoint (ALP < 1.67x ULN and TB ≤ ULN and ALP decrease from baseline ≥ 15% at week 52) and the three key secondary endpoints (Response to treatment based on ALP normalization at week 52 and Change from baseline through week 52 and through week 24 in PBC Worst Itch NRS) will be done for the following subgroups:

- Age at randomization (< 65 years, >= 65 years)
- Sex (Female, Male)
- Race (White, Others defined by American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander or Others))
- UDCA treatment at baseline (Yes/No)
- Prior OCA treatment (Yes/No)
- ALP level at baseline > 3x ULN (Yes/No)
- TB at baseline > ULN (Yes/No)
- TB at baseline > ULN or ALB at baseline < LLN (Yes/No)
- TB at baseline > 0.6x ULN (Yes/No)

- Geographic region: Europe, North America, Latin America, Other (including Turkey and South Africa)
- ALP > 3x ULN or TB > ULN at baseline (Yes/No). In case of mis-stratification during the randomization, the true screening value will be used.
- PBC Worst Itch NRS score ≥ 4 at baseline (averaged over the 14 days preceding randomization) (Yes/No). In case of mis-stratification during the randomization, the true screening value will be used.
- Cirrhotic defined by Liver stiffness at baseline ≥ 16.9 KPa at Fibroscan exam (Yes/No) and/or cirrhosis on histology
- Advanced disease stage defined as liver stiffness at baseline >10 kPa at Fibroscan exam and/or bridging fibrosis or cirrhosis on histology

4.2.8 Withdrawals, Dropouts, Loss to Follow-up

Discontinued patients who received the study drug are not to be replaced.

Over the DB period, follow-up of the patients who discontinue early the study drug will continue until the end of the DB period (V8 or until last V6 performed). The study procedures will be assessed as per Table 1 and Table 2.

To limit the occurrence of intercurrent events (ICEs) such as study treatment discontinuations and/or use of rescue medication such as OCA, the treatment allocation as well as values of ALP, GGT and 5'NT will remain blinded for the Investigator and for the patient up to the DB database lock.

Rescue therapy for PBC and Pruritus will be identified during the Blind Data Review Meetings (BDRM). Use of PBC rescue therapies and pruritus rescue therapies will be handled as ICE for the primary endpoint, first key secondary endpoint, and second and third key secondary endpoints respectively. The final list will be provided and approved at the last BDRM (i.e. before unblinding).

4.2.9 Missing, Unused, and Spurious Data

The patient's outcomes will be considered censored at the time of an ICE (study treatment discontinuation or use of rescue therapy) and any data observed one day or more after the occurrence of this event will be considered according to the estimand strategy for the main, sensitivity and supplementary analyses. However, these data will be included and flagged in any efficacy data listings. For any descriptive summary of efficacy (including the primary, the key secondary endpoints and all other secondary endpoints), these data will be described as collected.

In the summary tables of the safety data (laboratory, vital signs (including weight), physical examination, and ECG), any data observed more than 30 days after the last dose of study treatment intake for patients

who discontinued study treatment will be handled as missing data. However, this data will be included and flagged in any safety data listings.

For all the binary/categorical efficacy endpoints (including the primary endpoint, the first key secondary endpoint and all other binary secondary endpoints), the patient's outcome after the occurrence of an ICE (study treatment discontinuation or use of rescue therapy) will be considered as non-responders in a composite strategy for the main analysis. For the primary and key secondary endpoints, missing data at week 52 (i.e. visit 6) for patients without ICE will be replaced by the closest non-missing assessment from the DB treatment period for the main analysis, or patients will be considered as non-responders for the first supplementary analysis. CCI [redacted]

For continuous endpoints, the patient's outcome after the occurrence of an ICE (study treatment discontinuation or use of rescue therapy) will be considered as missing for the main analysis. Missing values for continuous efficacy endpoints with repeated measurements will be handled within the analysis itself via a MMRM with the assumption that the model specification will be correct, and that the data will be MAR.

CCI [redacted]

Unless otherwise specified, there will be no substitutions made to accommodate missing data points.
All data recorded during the DB period in the electronic Case Report Form (eCRF) or provided by external data transfer (ePRO and Central Laboratory data) will be included in data listings that will accompany the CSR.

Imputation of missing/partial AE dates will be done only to identify treatment-emergent AEs (TEAEs).

AE onset dates

- Partially missing onset dates will be imputed as follows:
  - When only Day is missing:
    - If Month & Year of the onset date are the same as Month & Year of the first administration date, the imputed onset date will be imputed as the minimum of the first administration date and the AE resolution date (imputed if needed).
    - Otherwise, the missing day will be replaced by "1."

CCI [redacted]

- When Day & Month are missing:
  - If Year of the onset date is the same as Year of the first administration date, the imputed onset date will be imputed as the minimum of the first administration date and the AE end date (imputed if needed).
  - Otherwise, the missing Day & Month will be replaced by "01 JAN."

- Completely missing onset dates for AEs will be imputed by the first administration date and the AE will be considered as treatment-emergent, unless the end date of the AE (imputed if needed) or the end year of the AE (if day and month are missing) is entered and is before the first administration date. If the end date is before the first administration date, the AE will not be considered as treatment-emergent and onset date of AE will be imputed by the end date of AE.

AE end dates

- If Day only is missing, incomplete end dates will be replaced by the last day of the month if it is not resulting in a date later than the date of the patient's death. In the latter case, the date of death will be used to impute the incomplete end date.
- If Day & Month are missing, Day & Month will be replaced by 31DEC if it is not resulting in a date later than the date of the patient's death. In the latter case, the date of death will be used to impute the incomplete end date.
- In all other cases the incomplete end date will not be imputed.
- If after imputation, the imputed AE start date is after the AE end date, the imputed AE start date will be replaced by the AE end date.

Treatment-emergent flag is defined according to the study treatment administered.

Treatment-emergent is defined as:

- All events which started after the first study treatment dose intake* (all episodes with start after or on (≥) the first study treatment dose intake) up to 30 days after the last study treatment dose intake (episodes with start before or on (≤) the last study treatment dose intake + 30 days) for the patients that discontinue the study treatment during DB period, and up to the date of the last DB data collection for the patients that complete DB period and switch in LTE, respectively.
- Or as all events which started before the first study treatment dose intake and worsened or became serious after or on the first study treatment dose intake (in case of multiple episodes of the same event before the first study treatment dose intake, the severity of the episode closest to the first study treatment dose intake will be considered) up to 30 days after the last study treatment dose intake for the patients that discontinue the study treatment during DB period, and up to the date of the last DB data collection for the patients that complete DB period and switch in LTE, respectively.

In case of missing severity, the emergence of AEs will be determined as shown in Table 4.

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

Table 4: Missing Severity and Definition of Treatment-emergent

Severity			Adverse event considered as
Closest episode before the first study treatment dose intake	Episode(s) after* the first study treatment dose intake (included)		
Missing	Missing	⇨	Emergent
Missing	Mild	⇨	Emergent
Missing	Moderate	⇨	Emergent
Missing	Severe	⇨	Emergent
Mild	Missing	⇨	Emergent
Moderate	Missing	⇨	Emergent
Severe	Missing	⇨	Non emergent

Concomitant medication and procedure dates imputation:

The same rules as for start and end dates of adverse events will be applied for partially or completely missing concomitant medications and procedures dates.

4.2.10 Visit Windows

For all analyses, the visit dates as collected in the eCRF will be used. However, the visits will be checked for accuracy according to protocol-specified intervals (Table 5) and sensitivity analyses may be deemed necessary considering this derivation, in particular for Week 52 comparative analyses.

For all efficacy and safety laboratory endpoints (from central laboratory), if the results are not available at the scheduled visit, then results from the closest unscheduled visit will be used if within the visit window and if still in DB treatment period.

Table 5: Evaluation Intervals for Efficacy and Safety Analysis

Evaluation	Protocol-Specified Interval
Screening	Day -84 to day -14
Baseline (V1) [1]	Day 1
Visit 2 (V2)	Day 29 (±7 days)
Visit 3 (V3)	Day 92 (±14 days)

CCI [redacted]

Visit 4 (V4)	Day 183 (±14 days)
Visit 5 (V5)	Day 274 (±14 days)
Visit 6 (V6)	Day 365 (±14 days)
Visit 7 (V7)	Day 547 (±14 days)
Visit 8 (V8)	Day 729 (±14 days)
LVDB	Within 91 days after last V5 (of the last patient) performed
Safety Contact – DB period	Day 456 and day 638 (±14 days)
Long term visit 1 (LT1)	91 days after V6/V8/LVDB (±14 days)
Long term visit 2 (LT2)	91 days after LT1 (±14 days)
Safety Contact – LTE period	Every 182 days, starting 91 days after LT2 (±14 days)
LT3 to LT11	Every 182 days, starting 182 days after LT2 (±14 days)
EOT/end of study (EOS) – Safety follow- up	16 to 30 days after last study drug intake

[1] Baseline value: Last non-missing assessment prior to first study drug intake

4.2.11 Data Derivation

PBC Worst Itch NRS average at Baseline will be calculated based on available data within 14 days prior the day of randomization (baseline period). The 4-week period post-baseline averages will be calculated based on available data every 28 days after the day of first study drug intake. The following formula for the PBC Worst Itch NRS average will be used:

[Sum (efficacy endpoint data during the period)/Number of Days with assessments recorded during the period]

For the baseline value, patients must have at least 4 values of the PBC Worst Itch NRS score during each of the two 7-day intervals prior to the day of randomization, otherwise the baseline value is considered as missing.

The 4-week period post baseline average NRS score will be calculated as follows:

- For a full 4-week period, at least 3 weeks out of 4 weeks complying with the rule of 4 out of 7 daily NRS score assessed and no less than 14 out of 28 daily (at least 50% of expected daily data) assessments of the full period will induce computation of the average 4-week period NRS score, otherwise, the average NRS score will be considered missing.
- For the last period (13th period) which is truncated by design, the rule will be flexible based on the number of expected weeks (2 to 4 weeks):
    - If two or three weeks are expected: each week has to be compliant with the rule of 4 out 7 daily

NRS score,
- If four weeks are expected: same rule as for a full 4-week period,

- Note: Data collected after W52 (365 days) will not be used

With a treatment duration of 52 weeks and calculated averages done every four weeks (28 days), 13 4-week periods average at most will be calculated.

PBC Worst Itch NRS average through week 52 and week 24 will be estimated based on the actual 4-week periods available for each patient, using the period 1 to period 13 and the period 1 to period 6 respectively.

For any combined response efficacy endpoint, e.g. primary endpoint, if one evaluated parameter reaches a non-responder condition (e.g. ALP >= 1.67 x ULN for the primary endpoint) while other parameters are missing then the patient will be considered as non-responder, otherwise the response will be considered as missing and handled as per analysis rule.

4.3 Interim Analyses

No interim analysis is planned for this study.

A full analysis will be performed after all patients have completed Week 52. Previous treatment allocation will remain blinded for the Investigator and the patient up to the DB database lock (see section 2.4).

4.4 Subject Disposition

All patients screened will be summarized overall and separately by geographical region and by country/site ID. Patients who failed screening and the primary reason for screening failure will be summarized.

All patients screened, randomized, and treated will be accounted for in this study. A table summarizing the number of patients per visit will be prepared.

Patient disposition will be tabulated and will include the number of patients treated, the number in each analysis population, the number who discontinued from treatment and reason(s) for discontinuation from treatment, and the number who discontinued from the study and reason(s) for discontinuation from study (including discontinuation due to Covid-19) prior to week 52 and prior to the end of the double-blind period.

Summary data will be presented by treatment group and overall. The denominator for all percentages is the total number of patients in each dose group. Treatment group is as randomized or as received according to the respective analysis population definition.

The following by-patient data listings will be presented:

- Study completion information including the reason for premature study discontinuation, if applicable.
- Inclusion/exclusion criteria.
- Patients inclusion in each of the analysis populations [ITT, pruritus ITT, PP, pruritus PP and Safety analysis sets].
- Patients excluded from the analysis sets (i.e. from the Safety, pruritus ITT, PP analysis set and pruritus PP set).
- Protocol Deviations.
- Patients that were unblinded before the DBL (who requested unblinding, date of unblinding, and the reason for unblinding).
- Patients affected by the COVID-19 related to study disruption

4.5 Demographic, Screening and Baseline Characteristics

All collected demographic, screening, baseline characteristics, prior medications, prior procedures and medical history data will be provided in data listings. The demographic, screening, baseline characteristics, prior medications, prior procedures and medical history data will also be summarized by treatment group and overall.

No formal statistical comparisons will be performed.

4.5.1 Demographics

Demographic characteristics at baseline (sex, age at randomization (years), age categories (<65 years/>=65 years), race, height (cm), weight (kg), Body Mass Index (BMI) (kg/m2)), vital signs (i.e., systolic blood pressure (mmHg), diastolic blood pressure (mmHg), heart rate (beats/minute) respiratory rate (breaths/minute) and temperature (C)) will be summarized using descriptive statistics for the ITT, the Safety and the PP, the exploratory, the pruritus ITT and the pruritus PP analysis sets.

4.5.2 Screening Characteristics

Screening characteristics data will be summarized for the ITT, for the Safety and for the PP analysis sets, as follows:

- Screening PGIS at Screening Visit 1.
- Screening laboratory measures of ALP, AST, ALT, GGT, CPK, total and direct bilirubin, ALB, creatinine, sodium, AFP, estimated glomerular filtration rate (eGFR),MELD-Na.
- Screening serum hematology measures of hemoglobin, hematocrit, WBC with differential, platelet count, PT, INR.
- Screening serology measures of HIV Ab I/II, Anti- HAV IgM, HBsAg, HCV Ab, HCV RNA.
- Screening urinalysis (dipstick) measures of specific gravity, pH, protein, glucose, ketones, bilirubin, urobilinogen, blood, nitrite, leukocytes.
- Screening urine-based β-hCG pregnancy test for all women of childbearingpotential.
- Screening serum pregnancy test (from central laboratory).

Note: for the screening laboratory parameters, the summaries will be provided by visit (i.e.: screening visit 1, screening visit 2, screening visit 3) where applicable.

Among the concomitant medications (see section 4.8.8), those which started during 30 days before and at first screening visit 1 including but not limited to UDCA, Colchicine, cholestyramine, rifampin, naltrexone, sertraline, statins, ezetimibe will be displayed by treatment group using Therapeutic Subgroup [ATC2] and Therapeutic Subgroup [ATC4] for the ITT and the Safety analysis sets.

Data collected at screening will be also included in the corresponding listings.

4.5.3 Baseline Disease Characteristics

Baseline disease characteristics will be summarized for the ITT, the Safety, the PP and the exploratory analysis sets, as follows:

- Duration since diagnosis of PBC in years defined as (year of randomization – year of diagnosis).
- Frequency and percentage of the randomization factor: ALP > 3x ULN or TB > ULN-(Yes/No)
- Baseline PBC Worst Itch NRS average over the 14 days preceding the randomization (V1).
- Baseline PBC Worst Itch NRS average score categorization ≥ 4 (Yes/No).
- Baseline laboratory measures of ALP, AST, ALT, GGT, CPK, total bilirubin, ALB, creatinine, creatine kinase, platelets, INR,MELD-Na.
- Frequency and percentage of baseline laboratory binary parameters (Yes/No): ALP level > 1.67x ULN, > 3x ULN, TB > ULN, ALB < LLN, TB > ULN or ALB < LLN, TB > 0.6x ULN.
- UDCA treatment at baseline
    - Yes/No
    - Duration (months): (randomization date – UDCA Initiation date + 1) / 30.4375
    - Duration (months) under stable dose: (randomization date – UDCA current date + 1) / 30.4375
    - Total daily dose (mg)
- Prior OCA treatment (Yes/No)
- Prior Bezafibrate treatment (Yes/No)
- Baseline Liver biopsy (Yes/No) (will be only displayed for the Exploratory analysis set)
- Baseline Transient Elastography (Fibroscan) (Yes/No)
- Baseline Liver stiffness (KPa)
- Baseline Liver stiffness (KPa) categorization ≥ 16.9 and/or cirrhosis on histology (Yes/No)
- Advanced disease stage defined as liver stiffness at baseline >10 kPa at Fibroscan exam and/or bridging fibrosis or cirrhosis on histology (Yes/No)

General medical history coded using MedDRA will be summarized per System Organ Class (SOC) and Preferred Term (PT) for the ITT and PP analysis sets. Medical history includes past and/or ongoing diseases or surgeries recorded in the Medical History eCRF. Reported terms will be coded using the latest MedDRA at the time of database lock.

4.5.4 Prior Medications

Any medication which was started before the first study drug intake will be counted as a prior medication.

Prior medications will be coded, summarized and listed for the ITT and Safety analysis sets in a similar way as described in section 4.8.8.

4.5.5 Prior Procedures

Any procedure that was started before the first study drug intake will be counted as a prior procedure. Prior procedures will be coded, summarized and listed for the ITT and Safety analysis sets in a similar way as described in section 4.8.9.

4.6 Efficacy Evaluation

4.6.1 Estimand Framework for Primary and Key Secondary Endpoints

The primary objective of the study is to evaluate the effect of Elafibranor (80 mg/day) on cholestasis over 52 weeks of treatment compared to placebo.

The key secondary objectives of the study are to evaluate the effect of Elafibranor (80 mg/day) over 52 weeks of treatment compared to placebo based on ALP normalization, and through 52 weeks and through 24 weeks on pruritus.

As detailed below, in line with ICH E9 (R1) addendum five attributes (treatment condition, population, endpoint, ICEs and population level summary) have been specified to translate the primary objective and key secondary objectives into treatment effect that is to be estimated (estimands):

- Primary estimand:

The **composite strategy** imputing non-response for patients who experienced ICEs prior to week 52 will be applied.

A. Treatment condition:	Administration of Elafibranor or Placebo on top of UDCA or in patients intolerant to UDCA.
B. Population:	Randomized patients with PBC and inadequate response or intolerance to UDCA.
C. Endpoint:	Response to treatment (binary variable) indicating a successful response at week 52 for a patient with ALP < 1.67x ULN, TB ≤ ULN and ALP decrease from baseline ≥15%, and who does not stop prematurely the study treatment nor uses any rescue therapy.

D. Intercurrent events:	To be considered as non-response irrespective of data after the study treatment discontinuation (ICE-1) or use of rescue therapy for PBC (ICE-2) being missing or not.
E. Population-level summary	Odds ratio between treatment groups.

The primary estimand can be defined as the odds ratio between treatment groups, from all randomized patients, achieving response at week 52 (defined as: ALP < 1.67x ULN, TB ≤ ULN and ALP decrease ≥ 15%), and not stopping prematurely the study treatment nor using rescue therapy for PBC.

In case of missing data at week 52 (i.e. visit 6) for patients without ICE, the closest non-missing assessment from the DB treatment period before or after the theoretical visit 6 date (i.e. visit 1 date + 52 * 7 days) will be taken into account. It may correspond to an unscheduled visit or a visit prior (i.e. visit 5) or after (i.e. visit 7) the visit 6.

CCI

[redacted]

CCI [redacted]

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

CCI [redacted]

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	Document Date:	10-May-2023

- First Key secondary estimand (ALP normalization)

The composite strategy imputing non-response for patients who experienced ICEs prior to week 52 will be applied.

A. Treatment condition:	Administration of Elafibranor or Placebo on top of UDCA or in patients intolerant to UDCA.
B. Population:	Randomized patients with PBC and inadequate response or intolerance to UDCA.
C. Endpoint:	Response to treatment (binary variable) indicating a successful normalization of ALP at week 52 without stopping prematurely the study treatment nor using any rescue therapy.
D. Intercurrent events:	To be considered as non-response irrespective of data after the study treatment discontinuation (ICE-1) or use of rescue therapy for PBC (ICE-2) being missing or not.
E. Population- level summary:	Odds ratio between treatment groups.

The first key secondary estimand can be defined as the odds ratio between treatment groups from all randomized patients who successfully normalize their ALP at week 52 and who do not prematurely stop the study treatment nor use any rescue therapy.

In addition, the supplementary analyses used for the primary estimand will be applied to the first key secondary estimand (see section 4.6.2.3).

- Second and third key secondary estimands (Change in pruritus through week 52 and week 24 respectively)

The hypothetical strategy assuming that patients who experienced an ICE would have continued within their treatment arm will be applied.

A. Treatment condition:	Administration of Elafibranor or Placebo on top of UDCA or in patients intolerant to UDCA.
B. Population:	Randomized PBC patients with baseline PBC Worst Itch NRS score ≥4 and inadequate response or intolerance to UDCA.
C. Endpoint:	Change from baseline through week 52 (and week 24) in 4-week period averaged PBC Worst Itch NRS respectively, (continuous).
D. Intercurrent events:	Any outcome value after a treatment discontinuation (ICE-1) or use of rescue therapy for pruritus (anti-pruritic medications, see section 4.2.8) (ICE-3) will be considered as missing.
E. Population-level summary:	Between treatment group difference in means of 4-week period averaged change in PBC Worst itch NRS through week 52 (and week 24) respectively.

The second and third key secondary estimands can be defined as the between treatment group difference in 4-week period means change from baseline from patients with baseline PBC Worst Itch NRS score ≥4 through week 52 and week 24 respectively, assuming they continued the assigned treatment after they experienced an ICE.

CCI [redacted]

In addition, the treatment policy strategy where all outcome values until week 52 and week 24 (respectively) will be used regardless of treatment discontinuation (ICE-1) or use of rescue therapy (ICE-3) (see section 4.6.2.3.3).

CCI [redacted]

CCI [redacted]

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	Document Date:	10-May-2023

CCI [redacted]

All efficacy analyses for primary and first key secondary endpoints will be based on the ITT analysis set and all efficacy analyses for second and third key secondary endpoints will be based on pruritus ITT analysis set. Descriptive statistics at each planned visit over the whole DB period (including V7 and V8 if applicable) will be summarized by treatment group (Elafibranor and Placebo) on efficacy endpoints (primary and secondary). CCI [redacted]

All the intercurrent events and their associated first date of occurrence will be included in by-patient data listing. Here the list of all the intercurrent events identified for the study:

- ICE-1: treatment discontinuation
  - ICE-1a: treatment discontinuation due to COVID-19 pandemic
  - ICE-1b: treatment discontinuation for any other reason
- ICE-2: use of rescue therapy for PBC
- ICE-3: use of rescue therapy for pruritus
- ICE-4: use of concomitant medication that creates or increases pruritus
- ICE-5: use of concomitant medication (including rescue therapy for PBC and pruritus) that reduces pruritus (anti-pruritus)

CCI [redacted]

4.6.2 Primary Efficacy endpoint analysis

4.6.2.1 Main Analysis

The efficacy of Elafibranor (80 mg/day) compared to placebo on cholestasis will be evaluated considering the response to treatment in terms of:

ALP < 1.67x ULN and TB ≤ ULN and ALP decrease from baseline ≥ 15% at week 52 (binary: YES/NO)

The null hypothesis for response to treatment based on the primary endpoint is that there is no difference in response rates between the Elafibranor and the placebo group. The alternative hypothesis is that there is a difference in response rates between both groups. The null hypothesis will be tested at a two-sided alpha of 0.05.

Patients who stopped prematurely the study treatment (ICE-1) or used rescue therapy for PBC (ICE-2) prior to week 52 assessment will be considered as non-responders.

In case of missing data at week 52 (i.e. visit 6) for patients without ICE (ICE-1 or ICE-2), the closest non-missing assessment from the DB treatment period before or after the theoretical visit 6 date (i.e. visit 1 date + 52 * 7 days) will be taken into account. It may correspond to an unscheduled visit or a visit prior (i.e. visit 5) or after (i.e. visit 7) the visit 6.

The number and percentage of patients with favorable response to treatment at the end of the 52 weeks of the DB period will be summarized by treatment group. The number and percentage of non-responder patients without ICE-1 nor ICE-2 as well as the number and percentage of non-responder patients due to ICE-1 or ICE-2 will be summarized. Response to treatment at Week 52 in patients with ICE (overall and by ICE-1 or ICE-2) will be described as well.

Finally, a descriptive summary by visit of cholestasis response using data of ALP and TB as collected will also be summarized in a separate table by treatment group.

The analysis of the primary efficacy endpoint will be conducted as follows:

The response proportions at week 52 will be compared between the treatment groups using the exact CMH test stratified by the randomization strata. The estimate of the odds ratio and the corresponding 95% exact CI and exact p-value will be provided. In addition, the risk difference between the treatment groups and 95% CI will be calculated using the Newcombe method stratified by randomization strata. For consistency, the Wilson score 95% CI for single proportion will be provided for within group description.

CCI [redacted]

CCI [redacted]

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	Document Date:	10-May-2023

CCI [redacted]

The treatment and response should be labelled so that the labels are alphabetically sorted in a given order (treatment: active treatment, placebo; response: event, no event).The main analysis will be conducted using the ITT analysis set CCI [redacted].
Figures of mean (+/- SE), mean change from baseline (+/- SE) and mean percentage change from baseline (+/- SE) over time up to week 52 and up to week 104 separately for ALP and TB will be produced using data as collected.

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

CCI

[redacted]

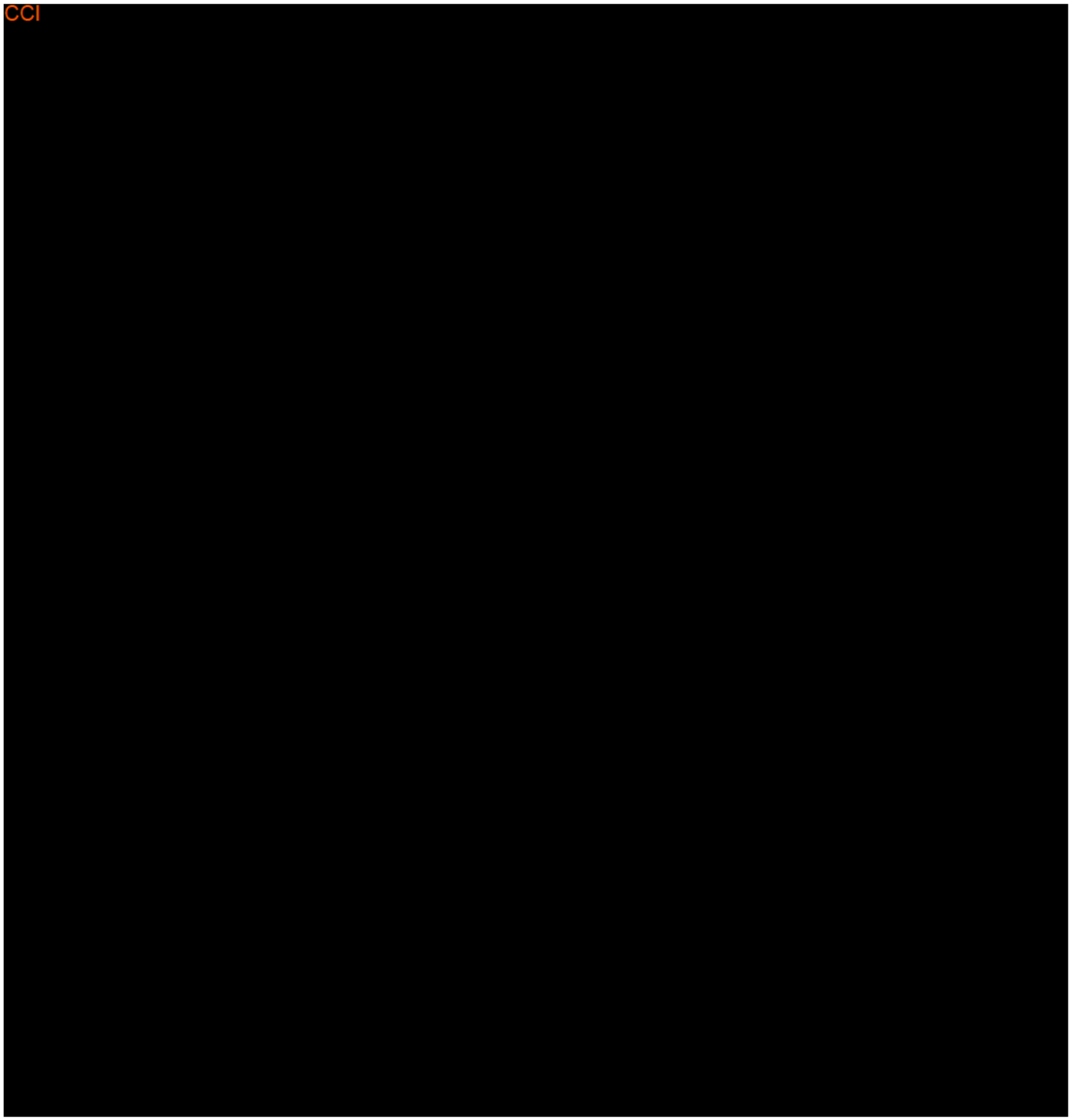
Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

CCI

CCI

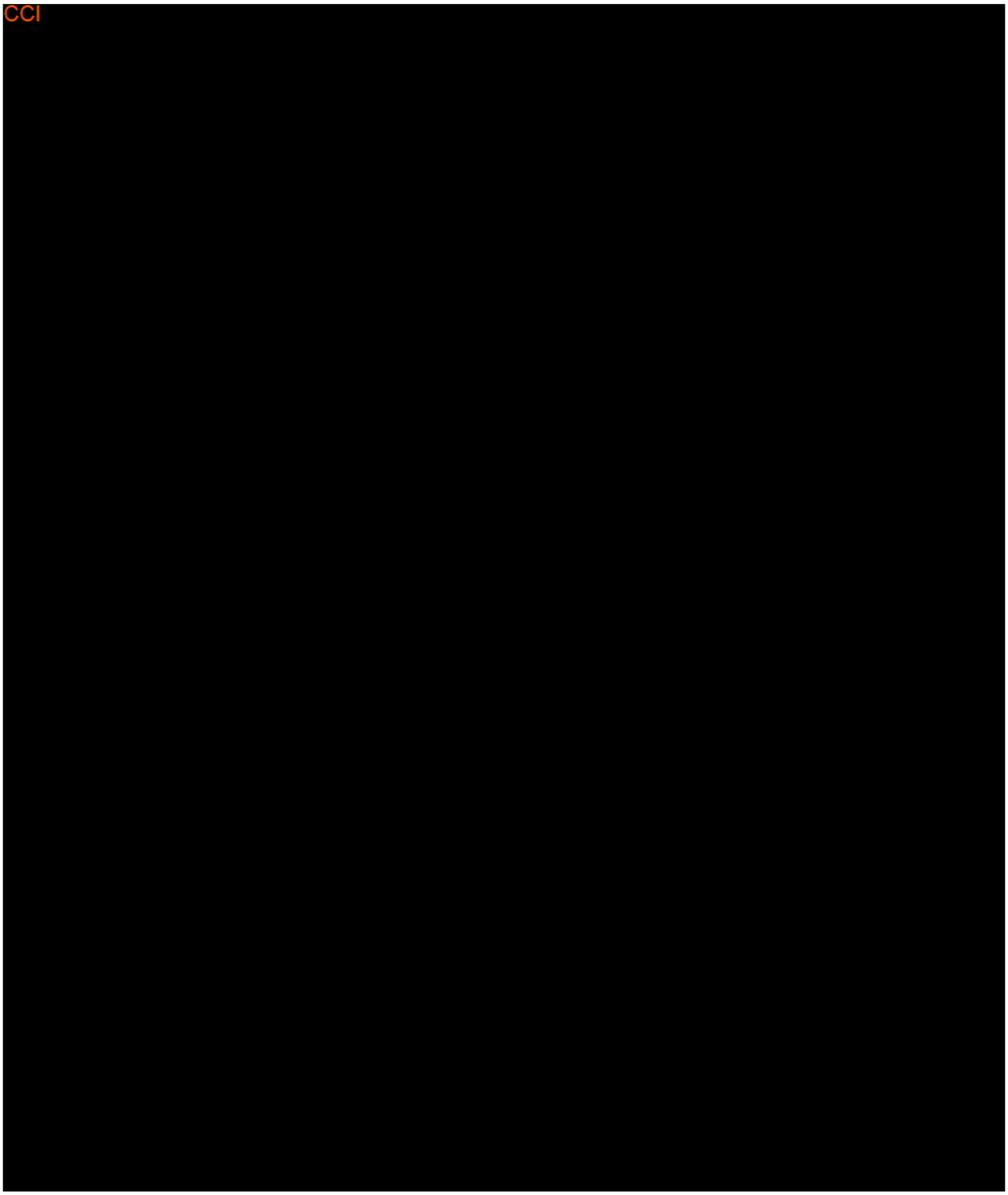
Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

CCI



Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

CCI



CCI

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

CCI

[redacted]

CCI [redacted]

Cytel **Statistical Analysis Plan**

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

CCI

[redacted]

CCI

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

CCI

[redacted]

CCI [redacted]

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

CCI

[redacted]

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

CCI

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

CCI

[redacted]

CCI [redacted]

4.6.3 First Key Secondary Efficacy Endpoint

The response to Elafibranor (80 mg/day) compared to placebo on cholestasis will be evaluated considering the response to treatment in terms of:

A successful normalization of ALP at week 52 (ALP ≤ ULN) - (binary: YES/NO).

The null hypothesis for response to treatment based on the first key secondary endpoint is that there is no difference in response rates between the Elafibranor and the placebo group. The alternative hypothesis is that there is a difference in response rates between both groups. The null hypothesis will be tested at a two-sided alpha of 0.05, only if the primary endpoint is statistically significant (see section 4.2.1).

Patients who stopped prematurely the study treatment (ICE-1) or took a rescue therapy for PBC (ICE-2) prior to week 52 assessment will be considered as non-responders.

In case of missing data at week 52 (i.e. visit 6) for patients without ICE (ICE-1 or ICE-2), the closest non-missing assessment from the DB treatment period before or after the theoretical visit 6 date (i.e. visit 1 date + 52 * 7 days) will be taken into account. It may correspond to an unscheduled visit or a visit prior (i.e. visit 5) or after (i.e. visit 7) the visit 6.

The number and percentage of patients with response to treatment at the end of the 52 weeks of the DB period will be summarized by treatment group, similar to the primary endpoint analysis.

The analysis of the first key secondary efficacy endpoint will be conducted similarly to the primary endpoint (see section 4.6.2). A sensitivity analysis to the statistical model, a supplementary analysis considering missing data at Week 52 in patients without ICE as non-response, a hypothetical strategy and a treatment policy strategy will be implemented the same way as for the primary efficacy endpoint. For the supplementary analyses with multiple imputation of the first key secondary endpoint, only the ALP will be imputed and not the TB (see section 4.6.2.3.1), re-running the imputation process with the same random

CCI [redacted]

seed including just the ALP variables.

A supplementary analysis to evaluate the impact of COVID-19 pandemic will be implemented the same way as for the primary efficacy endpoint.

4.6.4 Second Key Secondary Efficacy Endpoint

4.6.4.1 Main Analysis

For this endpoint "4-week period" is used as a synonym for the 28-day interval.

The response to Elafibranor (80 mg/day) compared to placebo on pruritus will be evaluated using the pruritus ITT analysis set and considering the response to treatment in terms of:

> Change in pruritus from baseline through week 52 on PBC Worst Itch NRS in patients with baseline PBC Worst Itch NRS score ≥4 (continuous)

The null hypothesis for response to treatment based on the second key secondary endpoint is that there is no difference in mean change from baseline through week 52 of PBC Worst Itch NRS score in patients with baseline PBC Worst Itch NRS score ≥4 between the Elafibranor and the Placebo group. The alternative hypothesis is that there is a difference in mean change from baseline through week 52 of PBC Worst Itch NRS score in patients with baseline PBC Worst Itch NRS score ≥4 between both groups. The null hypothesis will be tested at a two-sided alpha of 0.05, only if the primary endpoint and the first key secondary endpoint are statistically significant (see section 4.2.1).

PBC Worst Itch NRS scores for patients who stopped prematurely the study treatment (ICE-1) or took a rescue therapy for pruritus (ICE-3) prior to week 52 assessment will be considered as missing.

Mean PBC Worst Itch NRS values (13 4-week periods plus baseline) and change from baseline for each 4-week period (see section 4.2.11) will be summarized by treatment group (Elafibranor 80mg and Placebo). The number of compliant patients as well as the number of non-missing scoring days (see section 4.2.11) for baseline and for the 13 4-week periods will be provided by treatment group. Change from baseline at each 4-week period over time will also be plotted as well as the mean PBC Worst Itch NRS score over time using the least squares means (LSM) + 95% CI. Additionally, PBC Worst Itch NRS past week values at V7 and V8 as well as the change from baseline will be summarized only.

The analysis of the second key secondary efficacy endpoint will be conducted modeling the change from baseline values over the entire duration between baseline and week 52 via a MMRM adjusting for baseline PBC Worst Itch NRS score and the stratification factor [ALP > 3x ULN or TB > ULN-(Yes/No)] as possible covariates (see section 4.2.7). Please refer to the methods specified in section 4.6.7.1 for the verification of the dependency structure modelled if needed.

The estimated LSMeans, treatment differences, together with the 95% CIs and p value will be presented separately for the overall period (i.e. through week 52) and for each 4-week period until Week 52.

All 4-week periods [1,2, …, 13] until week 52 will be included as fixed effects along with treatment, treatment by 4-week period interaction and baseline PBC Worst Itch NRS average value. As a first intention,

the unstructured variance- covariance structure will be applied to model the within-patient variability. If there is no convergence, the following structures will be tested until the model succeeds to converge: heterogeneous Toeplitz structure (TOEPH), heterogeneous autoregressive(1) (ARH(1)), heterogeneous compound symmetry (CSH), Toeplitz structure (TOEP), autoregressive(1) (AR(1)), and compound symmetry.. The comparison between Elafibranor and Placebo will be based on the average (i.e. equal weights) of the 13 4-week period contrast.

Missing values will be handled within the analysis itself with the assumption that the model specification will be correct and that the data will be MAR (see section 4.2.9).

The 4-week periods will be considered as a repeated variable within a patient. The treatment by 4-week period interaction will be included in the model.

CCI

This main analysis will be repeated using the pruritus PP and ITT analysis sets.

CCI

CCI

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

CCI

[redacted]

CCI [redacted]

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

CCI

[redacted]

CCI [redacted]

Cytel **Statistical Analysis Plan**

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

CCI

[redacted]

CCI [redacted]

4.6.5 Third Key Secondary Efficacy Endpoint

As for the second key secondary endpoint, "4-week period" is used as a synonym for the 28-day interval.

The response to Elafibranor (80 mg/day) compared to placebo on pruritus will be evaluated using the pruritus ITT analysis set, repeated on ITT analysis set and considering the response to treatment in terms of:

> Change in pruritus from baseline through week 24 on PBC Worst Itch NRS score in patients with baseline PBC Worst Itch NRS score ≥4 (continuous)

The null hypothesis for response to treatment based on the third key secondary endpoint is that there is no difference in mean change of PBC Worst Itch NRS score in patients with baseline PBC Worst Itch NRS score ≥4 between the Elafibranor and the Placebo group through 24 weeks of treatment. The alternative hypothesis is that there is a difference in mean change of PBC Worst Itch NRS score in patients with baseline PBC Worst Itch NRS score ≥4 between both groups through 24 week of treatment. The null hypothesis will be tested at a two-sided alpha of 0.05, only if the primary endpoint, the first and the second key secondary endpoints are statistically significant (see section 4.2.1).

PBC Worst Itch NRS values for patients who stopped prematurely the study treatment (ICE-1) or took a rescue therapy for pruritus (ICE-3) prior to week 52 assessment will be considered as missing.

The analysis of the third key secondary efficacy endpoint will be conducted modeling the change from baseline values over the entire duration between baseline and week 52 via a MMRM adjusting for baseline PBC Worst Itch NRS score and the stratification factor [ALP > 3x ULN or TB > ULN-(Yes/No)] as possible covariates (see section 4.2.7). Please refer to the methods specified in section 4.6.7.1 for the verification of the dependency structure modelled if needed. All 4-week periods [1,2, …, 13] until week 52 will be included as fixed effects along with treatment, treatment by 4-week period interaction and baseline PBC Worst Itch NRS average value.

The estimated LSMeans, treatment differences, together with the 95% Cis and p value will be presented for the period through week 24 using contrast.

The same analyses, main and supplementary, as described above for the second key secondary endpoint will be repeated for the change from baseline in PBC Worst Itch NRS score through week 24 in patients with baseline PBC Worst Itch NRS score ≥4.

Note that all supplementary analysis will be provided using the pruritus ITT analysis set.

Cytel **Statistical Analysis Plan**

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

CCI

[redacted]

4.6.7 Other secondary endpoints

For each other secondary endpoint, descriptive summary by visit, for each post-baseline visit (i.e. Visit 2 – Week 4, Visit 3 – Week 13, Visit 4 – Week 26, Visit 5 – Week 39, Visit 6 – Week 52, Visit 7 – Week 78, Visit 8 – Week 104 if applicable) will be provided by treatment group on the ITT analysis set or pruritus ITT analysis set according to the endpoint and based on available data as collected (see section 4.2.9).

Analyses of all other secondary endpoints will be performed using the ITT analysis set or pruritus ITT analysis set according to the endpoint and using the composite strategy for binary endpoints (similar to the main analysis for the primary efficacy endpoint, except that missing data for patients without ICE will be kept missing), and hypothetical strategy for continuous endpoints (MAR assumption).

CCI [redacted]

4.6.7.1 Change from baseline in ALP at 4, 13, 26, 39 and 52 weeks (continuous)

Mean ALP values and change from baseline at each visit (baseline, V2, V3, V4, V5, V6, V7 and V8) will be summarized by treatment group (Elafibranor 80mg and Placebo) using the ITT analysis set. Change from baseline over time will also be plotted as well as the mean ALP over time using the LSM + 95% CI.

To assess the treatment effect on data collected up to week 52, change from baseline in ALP will be compared between treatment and placebo using the MMRM with stratification groups as fixed factors. Visit will be included as a fixed effect along with baseline ALP value. Visit will be considered as a repeated variable within a patient. The treatment by visit interaction will be included in the model.

Missing values for any reason will be handled within the analysis itself with the assumption that the model specification will be correct and that the data will be MAR (see section 4.2.9).

The treatment will be compared to placebo through week 52 for each scheduled visit from visit2-week 4 to visit 6-week 52. As a first intention, the unstructured variance- covariance structure will be applied to model the within-patient errors. If there is no convergence, the following structures will be tested until the model succeeds to converge: heterogeneous Toeplitz structure (TOEPH), heterogeneous autoregressive(1) (ARH(1)), heterogeneous compound symmetry (CSH), Toeplitz structure (TOEP), autoregressive(1) (AR(1)), and compound symmetry.

The model will be fitted using the Restricted Maximum Likelihood method (REML). Denominator degrees of freedom will be estimated using Kenward roger approximation. The statistical model will be used to estimate the mean treatment difference and its 95% CI.

The estimated LSM, treatment difference, together with the 95% CIs and p-value will be presented. Sample

CCI [redacted]

4.6.7.2 ALP response defined as 10%, 20% and 40% ALP reduction from baseline at week 52 (binary)

ALP response rates at 52 weeks will be classified as:

1) ≥ 10% decrease from baseline (Yes/No),
2) ≥ 20% decrease from baseline (Yes/No),
3) ≥ 40% decrease from baseline (Yes/No).

The analysis for each ALP response rates described above will be conducted similar to the main analysis for the primary efficacy endpoint (see section 4.6.2.1).

4.6.7.3 Response to treatment according to ALP < 1.5x ULN, ALP decrease from baseline ≥ 40% and TB ≤ ULN at week 52 (binary)

The analysis on this endpoint will be conducted the same way as the analysis described in section 4.6.2.1.

4.6.7.4 Response to treatment according to ALP < 3x ULN, AST <2x ULN and TB ≤ 1 mg/dL (Paris I) at week 52 (binary)

The analysis on this endpoint will be conducted the same way as the analysis described in section 4.6.2.1.

Note: the standard unit of Total bilirubin values is umol/L, then the conversion factor 1 mg/dL = 17.1 umol/L will be applied to derive the criterion.

4.6.7.5 Response to treatment according to ALP ≤ 1.5x ULN, AST ≤ 1.5x ULN and TB ≤ 1mg/dL (Paris II) at week 52 (binary)

The analysis on this endpoint will be conducted the same way as the analysis described in section 4.6.2.1

Note: the standard unit of Total bilirubin values is umol/L, then the conversion factor 1 mg/dL = 17.1 umol/L will be applied to derive the criterion.

4.6.7.6 Response to treatment according to TB decrease of 15% change from baseline at week 52 (binary)

The analysis on this endpoint will be conducted the same way as the analysis described in section 4.6.2.1.

4.6.7.7 Response to treatment according to normalization of abnormal TB (TB ≤ ULN) and/or ALB (ALB ≥ LLN) (Rotterdam) at week 52 (binary)

The endpoint will only be derived and analyzed in patients with abnormal total bilirubin or albumin at baseline.

Then patient will be considered as responder at week 52 only if total bilirubin and albumin are normal.

The analysis on this endpoint will be conducted the same way as the analysis described in section 4.6.2.1.

4.6.7.8 Response to treatment according to TB ≤ 0.6x ULN at week 52 (binary)

The analysis on this endpoint will be conducted the same way as the analysis described in section 4.6.2.1.

4.6.7.9 Response to treatment according to ALP ≤ 1.67x ULN and TB ≤ 1mg/dL (Momah/Lindor criterion, 2011) at week 52 (binary)

The analysis on this endpoint will be conducted the same way as the analysis described in section 4.6.2.1.

4.6.7.10 No worsening of TB defined as TB ≤ ULN at week 52 or no increase from baseline of more than 0.1x ULN at week 52 (binary)

The analysis on this endpoint will be conducted the same way as the analysis described in section 4.6.2.1.

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	Document Date:	10-May-2023

4.6.7.11 Complete biochemical response defined as normal ALP, TB, AST, ALT, albumin, and INR at week 52 (binary)

The analysis on this endpoint will be conducted the same way as the analysis described in section 4.6.2.1.

4.6.7.12 UK-PBC score at week 52 (continuous)

The continuous UK-PBC score will be calculated at baseline and from week 4 to week 52 (V1 to V6) in patients under UDCA through week 52 as follows:

UK-PBC score = 0.0287854*(alpEPxuln-1.722136304) - 0.0422873*(((altastEPxuln/10)^-1) - 8.675729006) + 1.4199 * (LN(bilEPxuln /10)+2.709607778) - 1.960303*(albxlln -1.17673001)-0.4161954*(pltxlln - 1.873564875)

Where

alpEPxuln=ALP /Upper Level Normal ALP at the timepoint

altastEPxuln=(ALT, AST or TA) /upper level normal of the value at the timepoint

bilEPxuln=bilirubin /upper level normal bilirubin at the timepoint

albxlln=alb /alb lower level normal at baseline

pltxlln=plt / plt lower level normal at baseline

The baseline survival function at the mean UK-PBC score S0(t) was: 0. 982, 0. 941, 0.893 at 5-, 10- and 15-year follow-up, respectively. The survival S(t) for any given patients was then calculated by S(t) = S0(t) exp(UK-PBC score).

Descriptive statistics, value and change from baseline, at baseline and from week 4 to week 52 (V1 to V6) will be provided on the continuous UK-PBC PBC score.

Descriptive statistics at baseline and from week 4 to week 52 (V1 to V6) will be provided on the continuous UK-PBC score predicted transplant-free survival at 5, 10 and 15 years.

The interpretation of this analysis should be made with caution. This analysis relies on strong assumptions that the score computed for patients under UDCA after one year of treatment is also applicable to patients with Elafibranor on top of UDCA, and the coefficients used to compute the score are applicable beyond 1 year of treatment.

4.6.7.13 GLOBE PBC score at week 52 (continuous)

The continuous GLOBE score will be calculated at the timepoints baseline and from week 4 to week 52 (V1 to V6) in patients under UDCA through week 52 as follows:

GLOBE score = 0.044378 * age at start of UDCA therapy + 0.93982 * LN(TBxuln) + 0.335648 * LN(ALPxuln)

– 2.266708 * ALBxlln – 0.002581 * platelet count per 10^9/L) + 1.216865.

Where

TBxuln = bilirubin/ULN at the timepoint

ALPxuln = ALP/ULN at the timepoint

ALBxlln = ALB/LLN at the timepoint

platelet count per 10^9/L = platelet count per 10^9/L at the timepoint

The baseline survival curve at the mean GLOBE score S0(t) was: 0.9385, 0.8429, 0.7361 at 5-, 10- and 15-year follow-up, respectively. The survival S(t) for any given patients was then calculated by S(t) = S0(t) exp(GLOBE score).

Descriptive statistics, value and change from baseline, at baseline and from week 4 to week 52 (V1 to V6) will be provided on the continuous GLOBE PBC score.

Descriptive statistics at baseline and from week 4 to week 52 (V1 to V6) will be provided on the continuous Globe PBC score predicted transplant-free survival at 5, 10 and 15 years.

The interpretation of this analysis should be made with caution. This analysis relies on strong assumptions that the score computed for patients under UDCA after one year of treatment is also applicable to patients with Elafibranor on top of UDCA, and the coefficients used to compute the score are applicable beyond 1 year of treatment.

4.6.7.14 Response to treatment based on the proportion of patients who normalized bilirubin (TB ≤ ULN) at week 52 (binary)

The endpoint will only be derived and analyzed in patients with abnormal total bilirubin at baseline.

The analysis on this endpoint will be conducted the same way as the analysis described in section 4.6.2.1.

4.6.7.15 Response to treatment based on the proportion of patients who normalized albumin (ALB ≥ LLN) at week 52 (binary)

The endpoint will only be derived and analyzed in patients with abnormal albumin at baseline.

The analysis on this endpoint will be conducted the same way as the analysis described in section 4.6.2.1.

4.6.7.16 Change from baseline to week 52 in hepatobiliary injury and liver function as measured by AST, ALT, GGT, 5' NT, total and conjugated bilirubin, albumin, INR and ALP fractionated (hepatic) (continuous)

The analysis on this endpoint will be conducted the same way as the analysis described in section 4.6.7.1.

Since the ALP fractionated (hepatic) is performed at baseline and Visit 6 - week 52 only, the MMRM reduces to an ANCOVA.

Mean ALP, GGT, AST and ALT values and relative change from baseline at each visit (baseline , V2, V3, V4, V5,V6, V7 and V8) will be summarized by treatment group (Elafibranor 80mg/day and Placebo).

The relative change ALP, GGT, AST, ALT from baseline to week 52 will be compared between Elafibranor and placebo using a nonparametric randomization-based Analysis of Covariance method (LaVange et al, 2005). This methodology uses weighted least squares on the treatment differences of outcome and covariate means. The resulting model estimates give the treatment effects for the outcomes, adjusting for the covariates added into the model.

The method will be applied using the SAS NParCov4 macro (©Zinc and Koch, 2001), adjusting for baseline ALP (or GGT, AST and ALT) level as a covariate.

Macro example for ALP relative change from baseline:

The macro will be applied to compare the relative change from baseline in ALP (PCTCHG) between the active treatment and placebo, adjusting for baseline ALP (BASE).

Assumptions for this model are minimal:

- Randomization to treatment
- Patients in the study are a simple random sample

CCI

[redacted]

CCI

4.6.7.17 Change from baseline to week 52 in biomarkers of inflammation as measured by hsCRP, fibrinogen, haptoglobin and TNF-α (continuous)

The analysis on this endpoint will be conducted the same way as the analysis described in section 4.6.7.1.

For the hsCRP, before computing the model, the hsCRP value will be log-transformed. After back transformation, the estimate of the between group geometric mean ratio and the corresponding 95% CI and p-value will be displayed.

4.6.7.18 Change from baseline to week 52 in immune response as measured by IgG and IgM (continuous)

The analysis on this endpoint will be conducted the same way as the analysis described in section 4.6.7.1.

4.6.7.19 Change from baseline to week 52 in biomarkers, non-invasive measures of hepatic fibrosis as measured by ELF test (HA, PIINP, TIMP-1), PAI-1, TGF-β, CK-18 (M65 and M30), Pro-C3 and liver stiffness measured by Transient Elastography (continuous)

The Enhanced Liver Fibrosis score (ELF) will be calculated based on the following parameters: Hyaluronic acid (HA) (ng/mL), PIIINP (ng/mL) and TIMP-1 (ng/mL) at baseline and each scheduled post-baseline visit.

The analysis on these measures will be conducted the same way as the analysis described in section 4.6.7.1.

4.6.7.20 Change from baseline to week 52 in lipid parameters as measured by TC, LDL-C, HDL-C, calculated VLDL-C and TG (continuous)

The analysis on these measures will be conducted the same way as the analysis described in section 4.6.7.1.

4.6.7.21 Change from baseline to week 52 in Fasting Plasma Glucose (continuous)

The analysis on these measures will be conducted the same way as the analysis described in section 4.6.7.1.

4.6.7.22 Change from baseline to week 52 in bile acids and biomarkers of bile acid synthesis as measured by bile acids, C4 and FGF-19 (continuous)

The analysis on these measures will be conducted the same way as the analysis described in section 4.6.7.1.

4.6.7.23 No worsening of pruritus based on PBC Worst Itch NRS through week 52 and through week 24 (binary)

A worsening is defined by a positive change in PBC Worst Itch NRS from baseline greater than 2. The criteria will be derived at week 52, i.e. at 4-week period 13 and at week 24, i.e. at 4-week period 6 respectively. It may also be repeated at other time points.
The analysis on this measure will be conducted in patients who can experience "worsening" (i.e. with PBC Worst Itch NRS score <8) at baseline and in the same way as the analysis described in section 4.6.2.1

4.6.7.24 Proportion of responders in PBC Worst Itch NRS score according to clinically meaningful change; at least 30% reduction; and one point, two points or three points decrease in score from baseline through week 52 and through week 24 in patients with baseline PBC Worst Itch NRS score ≥ 4

The proportion of responders in PBC Worst Itch NRS score at week 52, i.e. at 4-week period 13, and at week 24, i.e. at 4-week period 6, in patients with baseline PBC Worst Itch NRS score ≥ 4 (pruritus ITT analysis set) will be investigated in different ways as follows:

- According to clinically meaningful (clinical important response): a meaningful change analysis in PBC Worst Itch NRS score within the Psychometric Analysis, using scores on the PGIC and PGIS as anchors, will be performed when approximately 100 patients will attend week 52 (V6) visit. The meaningful change is defined by the amount of individual level change over a predefined period that could be interpreted as a meaningful benefit. This analysis will be performed before the database lock and in a blinded manner. The clinical important response (CIR), which represents the threshold for clinical relevance for a within-patient change, will be derived according to the determined threshold in the 4-week period PBC worst Itch NRS change from baseline.
- With at least a 30% of reduction in the 4-week period PBC Worst Itch NRS score change from baseline
- With at least a one-point decrease in the 4-week period PBC Worst Itch NRS score change from baseline
- With at least a two-point decrease in the 4-week period PBC Worst Itch NRS score change from baseline
- With at least a three-point decrease in the 4-week period PBC Worst Itch NRS score change from baseline

The analysis on these outcomes will be conducted the same way as the analysis described in section 4.6.2.1 and may also be repeated at other time points.

4.6.7.25 Proportion of participants achieving sustained improvement in PBC Worst Itch NRS (binary)

The proportion of participants achieving sustained improvement in PBC Worst Itch NRS score in patients with baseline PBC Worst Itch NRS score ≥ 4 (pruritus ITT analysis set) will be investigated in different ways as follows:

- With a clinically meaningful change (improvement) at all the last three 4-week periods (periods 11, 12 and 13)
- With at least a 30% reduction in the 4-week period PBC Worst Itch NRS score from baseline in all the last three 4-week periods (periods 11, 12 and 13)
- With at least a one-point decrease in the 4-week period PBC Worst Itch NRS score from baseline in all the last three 4-week periods (periods 11, 12 and 13)
- With at least a two-point decrease in the 4-week period PBC Worst Itch NRS score from baseline in all the last three 4-week periods (periods 11, 12 and 13)
- With at least a three-point decrease in the 4-week period PBC Worst Itch NRS score from baseline in all the last three 4-week periods (periods 11, 12 and 13)

The analysis on these outcomes will be conducted the same way as the analysis described in section 4.6.2.1.

Note that if only one period among periods 11, 12 and 13 is missing then the remaining 2 available periods will be used to derive the criteria and if more than 1 period among of the 3 periods, 11, 12 and 13 are missing the criteria will be defined as missing.

4.6.7.26 PGIC and PGIS

The PGIC and PGIS are both single item 5-point scales. Whereas the PGIC assesses the change in overall itch intensity since the baseline visit, the PGIS assesses the patient's impression of itch severity.

PGIC and PGIS are collected as anchors to facilitate the derivation of a clinically meaningful threshold. Both scales will be analyzed descriptively by visit and treatment groups. Shift from baseline for PGIS will also be presented. In particular, shift tables for patients who achieved one-category and two-category PGIS improvement respectively at visit 6 – week 52 will be presented.

4.6.7.27 Change from baseline to week 52 in 5-D Itch (continuous)

The items of the 5-D Itch Scale are grouped into five domains: duration, degree, direction, disability and distribution. The duration, degree and direction domains each include one item, while the disability domain has four items. All items of the first four domains are measured on a five-point Likert scale. The distribution domain includes 16 potential locations of itch, including 15 body part items and one point of contact with clothing or bandages.

Single-item domain scores (duration, degree and direction) are equal to the value indicated below the response choice (range 1–5).

The disability domain includes four items that assess the impact of itching on daily activities: sleep, leisure/social activities, housework/errands and work/school. The score for the disability domain is achieved by taking the highest score on any of the four items.

For the distribution domain, the number of affected body parts is tallied (potential sum 0–16) and the sum is sorted into five scoring bins: sum of 0–2 = score of 1, sum of 3–5 = score of 2, sum of 6–10 = score of 3, sum of 11–13 = score of 4, and sum of 14–16 = score of 5.

The scores of each of the five domains are achieved separately and then summed together to obtain a total 5-D score. 5-D scores can potentially range between 5 (no pruritus) and 25 (most severe pruritus).

Descriptive statistics will be provided by domain (considered as categorical) and for the Total 5-D score (considered as continuous) at each visit and by treatment group for the ITT and pruritus ITT analysis sets.

A comparative analysis between treatment groups on the Total 5-D score will be conducted the same way as the analysis described in section 4.6.7.1 using a MMRM with stratification groups as fixed factors. Visit will be included as a fixed effect along with baseline Total 5-D score. Missing values will be handled within the analysis itself with the assumption that the model specification will be correct and that the data will be MAR (see section 4.2.9). This analysis will be provided for the ITT and pruritus ITT analysis sets.

4.6.7.28 Change from baseline to week 52 in PROMIS Fatigue Short Form 7a (continuous)

In PROMIS Fatigue short form 7a, each of the 7 questions has five response options ranging in value from one (never) to five (always). First, the total raw score is computed by summing the values of the response to each question. All questions must be answered in order to produce a valid score using the scoring tables, otherwise the total raw score will be handled as missing. Finally, the score conversion table in APPENDIX C will be used to translate the total raw score into a T-score for each participant. The T-score rescales the raw score into a standardized score with a mean of 50 and a standard deviation (SD) of 10.

Descriptive statistics will be provided by question (considered as categorical), for the total raw score (considered as continuous) and the T-score (considered as continuous) at each visit by treatment group for the ITT and pruritus ITT analysis sets.

A comparative analysis between treatment groups on the T-score will be conducted the same way as the

analysis described in section 4.6.7.1 using a MMRM with stratification groups as fixed factors. Visit will be included as a fixed effect along with baseline T-score. Missing values will be handled within the analysis itself with the assumption that the model specification will be correct and that the data will be MAR (see section 4.2.9). This analysis will be provided for the ITT and pruritus ITT analysis sets.

4.6.7.29 Change from baseline to week 52 in Epworth Sleepiness Scale (continuous)

The Epworth Sleepiness Scale (ESS) is composed of 8 items investigating the same domain. Each item is scored using the 4-point Likert scale ranging from 0: *would never doze* to 3: *high chance of dozing*. The total score is the sum of all item scores and ranges from 0 to 24. If one or more item-scores are missing, the questionnaire is invalid, and the total score will be handled as missing.

The Total ESS score can be interpreted as follows: 0-5 Lower Normal Daytime Sleepiness; 6-10 Higher Normal Daytime Sleepiness; 11-12 Mild Excessive Daytime Sleepiness; 13-15 Moderate Excessive Daytime Sleepiness; 16-24 Severe Excessive Daytime Sleepiness.

Descriptive statistics will be provided by item (considered as categorical) and for the total score (considered as continuous) at each visit by treatment group for the ITT and pruritus ITT analysis sets.

A comparative analysis between treatment groups on the total score will be conducted the same way as the analysis described in section 4.6.7.1 using a MMRM with stratification groups as fixed factors. Visit will be included as a fixed effect along with baseline total score. Missing values will be handled within the analysis itself with the assumption that the model specification will be correct and that the data will be MAR (see section 4.2.9). This analysis will be provided for the ITT and pruritus ITT analysis sets.

4.6.7.30 Change from baseline to week 52 in PBC-40 (continuous)

In PBC-40 measures, each item is scored from 1 to 5 and the individual item scores are summed to give a total domain score (Symptoms, Itch, Fatigue, Cognition, Social, Emotional). Data should be considered by domain rather than in terms of a cumulative PBC-40 score. If data are missing from a domain (typically missed or duplicated answers) the whole domain should be discarded if <50% of items are completed. If >50% of responses are present, then the median value for the completed items in the domain should be ascribed to the missing item. The direction of scoring of some items is reversed for calculation of domain scores so that in all cases, high scores represent high impact and low scores low impact of PBC on quality of life.

Descriptive statistics will be provided by domain (Symptoms, Itch, Fatigue, Cognition, Social, Emotional) at each visit by treatment group for the ITT and pruritus ITT analysis sets.

A comparative analysis between treatment groups by domain (Symptoms, Itch, Fatigue, Cognition, Social, Emotional) will be conducted the same way as the analysis described in section 4.6.7.1 using a MMRM with stratification groups as fixed factors. Visit will be included as a fixed effect along with baseline domain score.

Missing values will be handled within the analysis itself with the assumption that the model specification will be correct and that the data will be MAR (see section 4.2.9). This analysis will be provided for the ITT and pruritus ITT analysis sets.

4.6.7.31 Change from Baseline to week 52 in health utility as measured by the EQ-5D-5L (continuous)

The EQ-5D-5L questionnaire is composed of 5 dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression), each one is scored from 1 (best) to 5 (worst), plus a visual analog scale numbered from 0 (worst) to 100 (best).

Descriptive statistics will be provided by dimension (considered as categorical) and for the VAS (considered as continuous) at each visit by treatment group for the ITT and pruritus ITT analysis sets.

A comparative analysis between treatment groups on the VAS will be conducted the same way as the analysis described in section 4.6.7.1 using a MMRM with stratification groups as fixed factors. Visit will be included as a fixed effect along with baseline VAS. Missing values will be handled within the analysis itself with the assumption that the model specification will be correct and that the data will be MAR (see section 4.2.9). This analysis will be provided for the ITT and pruritus ITT analysis sets.

4.6.7.32 Onset of clinical outcomes described as a composite endpoint of MELD-Na; Liver transplant; Uncontrolled ascites requiring treatment; Hospitalization for Variceal bleed, for Hepatic encephalopathy and for spontaneous bacterial peritonitis; death until week 52 (Binary)

Onset of clinical outcomes through week 52 will be described as a composite endpoint (binary) and will be composed of at least one of the following events:

- MELD-Na > 14 for patients with baseline MELD-Na≤12
- Liver transplant
- Uncontrolled ascites requiring treatment
- Hospitalization for new onset or recurrence of any of the following:
    a. Variceal bleed
    b. Hepatic encephalopathy defined as West-Haven/Conn of 2 or more
    c. Spontaneous bacterial peritonitis
- Death

Descriptive analyses of the composite endpoint and for each event will be conducted where number and percentage of patients will be summarized by treatment group.

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	Document Date:	10-May-2023

4.6.7.33 Change from baseline to week 52 in serum markers of bone turnover and in bone mineral density (hip and lumbar) assessed by DEXA scanning (continuous)

The analysis on the serum markers of bone turnover (CTX, P1INP), and the analysis on the bone mineral densities (in g/cm^2) of the total hip, femoral neck and the lumbar will be conducted the same way as the analysis described in section 4.6.7.1.

Since the DEXA scanning is performed at baseline and week 52 only, the MMRM reduces to an ANCOVA.

4.7 Pharmacokinetic Evaluations

The statistical considerations applicable to the popPK modeling, will be fully described in a dedicated Analysis Plan (popPK SAP).
All plasma concentration data will be summarized and listed. For participants from the Pharmacokinetics Analysis Set, individual plasma concentrations for elafibranor and GFT1007 will be summarized by timepoint and by visit, using descriptive statistics for continuous variables (number of available observations, number of below the limit of quantification, mean, SD, % coefficient of variation (CV), geometric mean and %CV geomean, median, minimum, maximum).

4.8 Safety Analyses

The analysis of the extent of exposure and compliance to treatment will be conducted using theSafety analysis sets. Safety analyses of AEs, laboratory parameters, and vital signs will be conducted using the Safety analysis set. Analyses of concomitant medications and concomitant procedures will be conducted using the ITT analysis set and will be repeated on the Safety analysis set only if safety analysis set is different from ITT.

4.8.1 Extent of Exposure and Compliance

4.8.1.1 Extent of Exposure

Duration of treatment and duration of exposure (in weeks) from Study Visit 1 to Week 52 visit and until the end of double-blind period (EODB) will be summarized by treatment group (see section 4.2.2). The number and percentage of patients in categories defined by duration of exposure in weeks will also be provided. The definition is as follows:

- Total duration of exposure (in weeks):
  - Up to week 52: (minimum of (date of last dose, date of death, date of data cutoff, date of week 52 visit) – first dose date + 1)/7
  - Up to EODB period: (minimum of (date of last dose, date of death, date of data cutoff, date of last visit during the double-blind period) – first dose date + 1)/7
- Total duration of exposure (in weeks) up to week 52 and up to the EODB period in categories: ≥0 to <12, ≥12 to <24, ≥24 to <36, ≥36 to <48, ≥48 to <60, ≥60 to <72, ≥72 to <96, ≥96 to <104, >= 104

Exposure data will be listed in patient data listing.

4.8.1.2 Compliance

The patient should take one tablet of Elafibranor (80 mg/day) or placebo every day:

While the patient is being treated with the study drug, the patient will be directed to bring back all used and unused cartons and blisters at each visit. Compliance will be checked by the Investigator during those visits and recorded on the eCRF. Percentage compliance will be calculated as:

100 * tablets taken/expected tablets taken,

Where the tablets taken is defined as the total number of tablets dispensed – the total number of tablets returned. It is assumed that tablets not returned have been taken by the patient. If not the case, it may lead to overestimation of the compliance.

The expected tablets taken is defined as the total duration of exposure in days.

The number and percentage of compliant patients will be presented for the Safety analysis set, where compliant is defined as percentage compliance between 80.0% and 120.0% inclusive. The following percentage compliance categories will be presented :<80.0%; ≥80.0% to ≤120.0%; >120.0%.

Compliance data will be summarized until Week 52 visit and until the EODB period. Cumulative dose (mg) and total expected dose (mg) until Week 52 visit and until the EODB period will also be summarized for Elafibranor arm.

Compliance data will also be listed.

4.8.2 Adverse Events

AEs will be coded using MedDRA and displayed in tables and listings using System organ class (SOC) and grouped Preferred term (PT). Some groupings of Preferred Terms of AEs will be done. The intent of grouping PTs for safety analyses is to avoid splitting of PTs to ensure that adverse events are adequately captured. Two different approaches will be used to group PT: the first will be to use a three-level coding, system organ class (SOC), high level term (HLT) and preferred term (PT), the second will be to use specific SMQ provided by the medical team before DBL.

Analyses of AEs will be performed for those events that are considered treatment-emergent AE (TEAE), where treatment-emergent is defined as any AE with onset on or after the date of first administration of study treatment and up to the date of the last DB data collection for the patients that complete DB period and will switch in LTE and up to 30 days after the date of the last study treatment for the patients that discontinue the study treatment during DB period, respectively or any event with start date prior to first dose of treatment whose severity worsened in intensity on or after the date of first dose of study treatment and up to the date of the last DB data collection for the patients that complete DB period and switch in LTE and up to 30 days after the date of the last study treatment for the patients that discontinue the study treatment during DB period, respectively.

An overview of AEs will be provided displaying the number of events along with the number and percentage

of patients with at least one:

- AE
- AESI
- Serious AE
- Severe AE
- TEAE
- TEAE by maximum severity
- TEAE related to study medication
- Serious TEAE
- Serious TEAE related to study medication
- TEAE leading to treatment discontinuation
- TEAE leading to study discontinuation
- TEAE related to study medication leading to treatment discontinuation
- TEAE related to study medication leading to study discontinuation
- Serious TEAE leading to treatment discontinuation
- Serious TEAE related to study medication leading to treatmentdiscontinuation
- Serious TEAE leading to study discontinuation
- Serious TEAE related to study medication leading to study discontinuation
- Serious TEAE leading to death
- Serious TEAE related to study medication leading to death.
- TEAE related to study procedure

AEs are summarized giving the total number of events as well as by patient incidence rates, therefore, in any tabulation, a patient contributes only once to the count for a given AE (SOC or 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. Tables will be sorted by descending frequency of SOC/PT or SOC/HLT/PT (if applicable) in Elafibranor group.

The number of events along with the number and percentage of patients with any AE will be summarized by treatment group and by SOC/PT.

The number of events along with the number and percentage of patients with any TEAE will be summarized by treatment group. Three distinct tables will be provided by SOC/PT, by SOC/HLT/PT, and by SMQ provided by the medical.

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	Document Date:	10-May-2023

TEAEs with frequency of patients > 1%, >5% and >10% in any treatment group will also be summarized by SOC/PT and by SOC/HLT/PT.

The number of events along with the number and percentage of patients with any severe TEAE will be summarized by treatment group and by SOC/PT.

The number of events along with the number and percentage of patients with any TEAE by maximum severity (mild, moderate, severe) and TEAE related to study medication (i.e. related, possibly related, and not assessable) will be summarized by treatment group and by SOC/PT. Relationship to study medication is not expected to be missing after data cleaning; however, if relationship is confirmed as missing, the AE will be considered as treatment related.

The number of events along with the number and percentage of patients with any serious TEAE and with any serious TEAE related to study medication will be summarized by treatment group and separately by SOC/PT and SOC/HLT/PT.

The number of events along with the number and percentage of patients with any TEAE leading to treatment discontinuation will be summarized by treatment group and separately by SOC/PT and SOC/HLT/PT.

The number of events along with the number and percentage of patients with any TEAE related to study medication leading to treatment discontinuation will be summarized by treatment group and by SOC/PT.

The number of events along with the number and percentage of patients with any TEAE leading to study discontinuation and with any TEAE related to study medication leading to study discontinuation will be summarized by treatment group and by SOC/PT.

The number of events along with the number and percentage of patients with any serious TEAE leading to treatment discontinuation and with any serious TEAE related to study medication leading to treatment discontinuation will be summarized by treatment group and by SOC/PT.

The number of events along with the number and percentage of patients with any serious TEAE leading to study discontinuation and with any serious TEAE related to study medication leading to study discontinuation will be summarized by treatment group and by SOC/PT.

The number of events along with the number of events along with the number and percentage of patients with any serious TEAE leading to death and with any serious TEAE related to study medication leading to death will be summarized by treatment group and by SOC/PT.

The number of events along with the number and percentage of patients with any TEAE related to the study procedure will be summarized by treatment group and by SOC/PT. In these tabulations, each patient will contribute only once (i.e., the most related occurrence or the most intense occurrence) to each of the incidence rates in the descriptive analysis, regardless of the number of episodes.

No formal hypothesis-testing analysis of AEs incidence rates will be performed.

All AEs (more especially, death, serious TEAEs, non-treatment-emergent SAEs, TEAEs leading to withdrawal or temporary withdrawal of study drug) occurring on study will be listed in patient data listings.

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 correspondingto:
  - Muscle pain or Myalgia
  - Muscle spasms or Tremor
  - Muscle weakness
- Transaminases elevations from baseline of severe intensity or leading to permanent study drug discontinuation
- Autoimmune hepatitis
- Liver injury events of severe intensity corresponding to:
  - Hepatic injury
  - Hepatic impairment
  - Hepatic failure
- Gastrointestinal symptoms of severe intensity correspondingto:
  - Abdominal pain
  - Constipation
  - Diarrhea
  - Nausea
  - Decreased appetite
  - Vomiting
  - Acute cholecystitis
  - 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 correspondingto:
  - Renal injury
  - Renal failure
  - Renal impairment
  - Renal colic
- Neurological abnormalities of moderate to severe intensity corresponding to:
  - Tremor
  - Ataxia
  - Fasciculations
- Parkinson's Disease
- Peripheral edema of moderate to severe intensity
- Weight gain of more than 5% from baseline
- Major Adverse Cardiovascular Events corresponding to:
  - Non-fatal myocardial infarction/unstable angina
  - Non-fatal stroke

  - Unstable Angina
  - Hospitalization for Heart Failure
  - Coronary Revascularization (bypass or percutaneous coronary intervention)
- 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.

A between-group comparison will be performed using the exposure-adjusted incidence rates (EAIR) of treatment-emergent AEs (TEAE), serious TEAE, TEAE leading to treatment discontinuation, TEAE leading to study discontinuation and AESIs.
The exposure-adjusted incidence rate (patients per patient-year) is defined as:

Number of patients with at least one event / Total exposure time for all patients at risk

and will be calculated by treatment arm.

The exposure time will be differentiated according to the occurrence or not of the event and will be summarized. If the patient experiences an event: their exposure time will be the time until initial occurrence of their events. If patient does not experience an event:

- For patients who discontinue the study treatment during the DB period, their exposure time will be censored at the time of the last treatment intake date plus 30 days (minimum of ((last treatment date + 30 days, date of death, database cut-off date) – first treatment date + 1)/365.25). The addition of 30 days matches the maximum time for scheduling the safety follow-up visit and the SAE reporting period after treatment discontinuation.
- For patients who don't discontinue the study treatment during the DB period, exposure time will be censored on the patient DB period end date (minimum date of the last double-blind visit, first

LTE treatment intake (if complete) – 1). The addition of 30 days after last dose of DB treatment does not apply to those patients that will complete DB period and will enter in the LTE period to receive open label elafibranor.

The comparison will be summarized using the EAIR difference, which will be accompanied by a 95% CI for the overview of adverse event and AESI, TEAE, serious TEAE, TEAE leading to treatment discontinuation, TEAE leading to study discontinuation. The 95% CI for the EAIR difference will be computed using the He and al method (He et al, 2015).

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

The number and percentage of patients who developed lipid abnormality and/or dysglycemia through the DB period will be summarized by SMQ code/PT and by treatment group using the following SMQ:

- Hypoglycaemia (SMQ Code 20000226)
- Hyperglycaemia/new onset diabetes mellitus (SMQ Code 20000041)
- Dyslipidaemia (SMQ Code 20000026)

The risk difference and its 95%CI (Newcombe) between Elafibranor and Placebo will also be provided.

The number and percentage of patients who developed hepatic injury and hepatic failure through the DB period will be summarized by custom queries/PT and by treatment group using following custom queries presented in Appendix F. These custom queries may be updated at the last BDRM before unblinding.

The risk difference and its 95%CI (Newcombe) between Elafibranor and Placebo will also be provided.
The incidence rate per patient exposure years, the EAIR difference and its 95% CI will be computed using the He and al method (He et al, 2015) by treatment group and according to the custom queries will be provided.

Time to the onset of first TEAEs, defined as (start date of the first TEAE – first study treatment date +1)/7, and duration of first TEAEs, defined as (end date of first TEAE – start date of first TEAE +1)/7 and for ongoing TEAE, the end date will be imputed as the datacut date, will be summarized by preferred term and by treatment group. The table will be sorted by descending frequency of PT in Elafibranor group.

4.8.3 Adjudicated events

The number of events along with the number of events along with the number and percentage of patients with any adjudicated adverse events including DILI will be summarized by treatment group. Only positively adjudicated events will be displayed. The number and percentage of the adjudicated outcome of Clinical Outcomes and DILI will be also summarized by treatment group. If the CEC identified any un-reported events, they will be summarized by treatment group.

All adjudicated events for each patient will be provided in data listings.

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	Document Date:	10-May-2023

Note: for the adjudications of events, only collected data in the EDC will be used.

4.8.4 Laboratory Data

Clinical laboratory evaluations include hematology, chemistry and urinalysis. Clinical laboratory values will be expressed using standard international (SI) and conventional (US) units with normal ranges and respective units for each parameter provided. All analyses will be done on the central laboratory data including baseline derivation. Local laboratory data will be provided on a patient data listing.

In the event of repeated post-baseline values, the first non-missing value per study day/time will be used. In the following calculations, observed data more than 30 days after study treatment discontinuations for patients who discontinued study treatment are handled as missing values (see section 4.2.9), but are included and flagged in any data listings.

Potentially clinically significant laboratory abnormalities (PCSA) will be determined based on the CTCAE criteria version 5.0 and defined as Grade 2 or higher for the laboratory tests of interest (see section 10). Note that CTCAE scale could not cover all data cases and for such cases, clinically relevant PCSA thresholds have been added in Appendix D (see section 10).

Serum hematology and serum chemistry (see Table 2) assessed from blood and urine samples taken at baseline, V2, V3, V4, V5 and V6 will be summarized by treatment group. If applicable, values obtained at V7 and V8 will also be summarized by treatment group. Changes from baseline in serum hematology and chemistry parameters will be summarized by treatment group and scheduled visit using shift tables to present the frequencies and percentage of patients below, within and above the normal ranges at baseline and at each post-baseline visit. In addition, change from baseline to worst value during the DB period in serum hematology and chemistry parameters will be summarized by treatment group. Worst value and change from baseline through the BD period, including scheduled and unscheduled visits, will be classified by the medical team into the following categories: high worst, low worst or both (high and low worst) according to the parameter (see Appendix E section 11).
Figures of chemistry mean creatinine and mean creatine phosphokinase will be produced using data as collected for the change from baseline over time and separately up to week 52 and up to week 104.

A summary descriptive, shift table and a patient data listing of potentially clinically significant laboratory abnormalities for the laboratory tests of interest (see section 10) identified using the CTCAE v5 will be provided by treatment group, by scheduled visit and through DB period.

For serum creatinine, also the percentage change from Baseline by visit and the high worst value and percentage change from baseline through DB period (including scheduled and unscheduled visits) will be presented and all serum creatinine values of patients with a post-baseline percentage change increase by 25% or greater will be listed.

For urine ACR, the percentage change from Baseline by visit and the high worst value and percentage change from baseline through DB period (including scheduled and unscheduled visits) will be presented.

For patients with (at least one) increase by 25% change from baseline in serum creatinine, the percentage change from baseline by visit and the high worst value and percentage change from baseline through DB period (including scheduled and unscheduled visits) will be repeated for urine ACR.

For Cystatin C, the percentage change from Baseline will be presented for all patients by visit and the high worst value and percentage change from baseline through DB period (including scheduled and unscheduled visits), and in the subgroup of patients with (at least one) increase by 25% change from baseline in serum creatinine. For patients with (at least one) increase by 25% change from baseline in serum creatinine, summary statistics by visit and the high worst value and percentage change from baseline through DB period (including scheduled and unscheduled visits) for Cystatin C will be repeated.

For eGFR, the CKD-EPI - Cystatin C parameter will be derived (Inker et al. 2012), the CKD-EPI creatinine equation and the MDRD parameters will come from central laboratory. The percentage change from baseline will be presented separately for those three eGFR parameters by visit in the subgroup of patients with (at least one) increase by 25% change from baseline in serum creatinine and the high worst value and percentage change from baseline through DB period (including scheduled and unscheduled visits) will also be displayed.

Biobank samples (see Table 2) collected from additional blood samples from patients, who have given their consent, at baseline, V2, V3, V4, V5 and V6 - and if applicable at V7 and V8- will be summarized by treatment group.

Frequency and percentage of changes from Baseline classified as increase by 100%, 200% and 300% or greater in biomarker results will be given (including other biochemical safety markers (Cystatin C, urine ACR, urinary myoglobin) and biomarkers of inflammation (HsCRP, fibrinogen, haptoglobin, TNF-α, IL-6, ELF (HA, PIINP, TIMP-1), PAI-1, TGF-β, CK-18 (M65 and M30), Pro-C3).

Evaluation of Drug-Induced Serious Hepatotoxicity (eDISH) plots will present the concomitant values of aminotransferase (AT: ALT or AST) and TB at the visit corresponding to the peak of AT.
Distinct plots will be produced separately with regards to the peaks of ALT and AST. For each patient and for both ALT and AST, the strategy is as follows:

- Among all the visits (includes only values as described in the beginning of section 4.2.9), identify the maximum value of AT
- If the peak value occurs at just one visit, keep the value of TB at the visit corresponding to the peak of AT
- If the peak value occurs at more than one visit, keep the highest value of TB at any of the visits corresponding to the peak of AT
- If the peak value occurs at a visit/visits with no corresponding value/values of TB, 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 TB at any of the visits equidistant from the corresponding visit/visits of peak AT
- Classify patients into two subgroups: baseline value of ALT ≤ ULN or baseline value of ALT > ULN.
    - If baseline value of ALT ≤ ULN, express TB in xULN on the y-axis and AT in xULN on the x-axis. Add horizontal and vertical lines corresponding to TB = 2 ULN and AT = 3 ULN,

respectively.

- If baseline value of ALT > ULN, present two plots:
  - Express TB in xBaseline on the y-axis and AT in xBaseline on the x-axis. Add horizontal and vertical lines corresponding to TB = 2xBaseline and AT = 3xBaseline, respectively. For TB values > 2xBaseline, values identified as > 2 ULN will be plotted differently than values ≤ 2 ULN.
  - Express TB in xULN on the y-axis and AT in xBaseline on the x-axis. Add horizontal and vertical lines corresponding to TB = 2 ULN and AT = 3xBaseline, respectively.

Increases in serum creatinine through the end of the 52-week DB period and through the last post-baseline visit during the double-blind period will be presented as:

- Occurrence of at least one post baseline value > ULN (only on the subset of patients with baseline value ≤ ULN)
- 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:
  - ≥ 1.5xBaseline and < 2.0xBaseline, or ≥ 0.3 mg/dL increase
  - ≥ 2.0xBaseline and < 3.0xBaseline
  - ≥ 3.0xBaseline or ≥ 4 mg/dL.

Hepatic laboratory test elevations (HLTE) will be based on fold changes above the upper limit of normal (ULN).

Cumulative elevations

- ALT value > 1x, > 3x, > 5x, > 10x, > 20x ULN and ALT change from baseline >250 U/L
- AST value > 1x, > 3x, > 5x, > 10x, > 20x ULN and AST change from baseline >250 U/L
- ALP value > 2x, > 3x ULN and ALP change from baseline >150 U/L
- Total bilirubin value > 2x, > 5x, > 8x ULN and TB change from baseline > 2 mg/dL
- Direct bilirubin value > 2x, > 5x ULN and Direct bilirubin change from baseline > 1 mg/dL
- GGT value > 2x ULN
- INR value > 1.5x, 3x and 5x ULN

These are based on the highest ratio of value in the double-blind safety period, including scheduled and unscheduled visits, to ULN for each laboratory test. Note that patients may be in > 1 category for each laboratory test.

HLTE will be displayed in frequency tables presenting the number and percentage of patients in categories and the associated risk difference and its 95% CI (Newcombe).

The incidence rate per patient exposure years, the EAIR difference and its 95% CI will be computed using

the He and al method (He et al, 2015) by treatment group and according to the HLTE will be provided.

In order to identify patients who meet potential DILI triggers (parameters being assessed at the same visit), the following biochemical triggers will be summarized:

- For patients with all normal AST, ALT, ALP, TB, direct bilirubin and GGT at baseline:
  - (ALT >= 3x ULN or AST >= 3x ULN) and TB >= 2x ULN
  - ALT >= 5x ULN or AST >= 5x ULN
  - ALT >= 8x ULN or AST >= 8x ULN
  - (ALT >= 3x ULN or AST >= 3x ULN) and INR > 1.5
  - ALP >= 2x ULN and direct bilirubin >= 2x ULN
  - GGT >= 2x ULN and direct bilirubin >= 2x ULN
- For patients with at least one abnormal laboratory parameter (AST, ALT, ALP, TB, direct bilirubin or GGT) at baseline
  - (ALT >= 3x baseline or AST >= 3x baseline) and TB >= 2x BL
  - ALT >= 3x baseline or AST >= 3x baseline
  - ALT >= 5x baseline or AST >= 5x baseline
  - (ALT >= 2x baseline or AST >= 2x baseline) and INR > 1.5
  - ALP >= 2x baseline and ALP > ULN and direct bilirubin >= 2x baseline and direct bilirubin > ULN
  - GGT >= 2x baseline and GGT > ULN and direct bilirubin >= 2x baseline and direct bilirubin > ULN

Potential DILI triggers will be displayed in frequency tables presenting the number and percentage of patients in categories and the associated risk difference and its 95%CI.
The incidence rate per patient exposure years (see section 4.8.1.1 for derivation rules), the EAIR difference and its 95% CI will be computed using the He and al method (He et al, 2015) by treatment group and according to the Potential DILI triggers will be provided.

Serum bile acids and biomarkers of bile acid synthesis, biomarkers of hepatic fibrosis and/or inflammation, additional kidney safety markers and immunoglobulins (see Table 1 and Table 2) at baseline and each scheduled post-baseline visit of the DB period will be summarized by treatment group.

All laboratory data will be provided in data listings.

4.8.5 Transient elastography

In the following calculations, observed data after the occurrence of the ICEs start of rescue therapy or more than 30 days after other study treatment discontinuations for patients who discontinued study treatment are handled as missing values (see section 4.2.9), but are included and flagged in any data listings.

The actual value and change from Baseline to each scheduled visit evaluation (baseline, V4, V6, and if applicable V7 and V8) for transient elastography (continuous, including Stiffness / Liver Stiffness Measurement in kPa, Interquartile range (IQR) in kPa and Interquartile range (IQR)/median in %) will be

summarized by treatment group.

By-patient listings of transient elastography measurements will be presented in data listings.

4.8.6 Vital Signs and Physical Examinations

In the following calculations, observed data more than 30 days after study treatment discontinuations for patients who discontinued study treatment are handled as missing values (see section 4.2.9), but are included and flagged in any data listings.

The actual value and change from Baseline to each study visit evaluation (baseline, V2, V3, V4, V5, V6, and if applicable V7 and V8) for vital signs (continuous, including weight and derived BMI using height at screening) will be summarized by treatment group. In addition, temperature, systolic and diastolic blood pressure, respiratory and heart rates, will be summarized descriptively, including the number and percent of patients with normal, abnormal, and clinically significant abnormal results at each applicable scheduled visit and the worst value through DB period (including scheduled and unscheduled visits) by treatment group; shifts from Baseline in those parameters to each scheduled visit and through DB period will also be presented.

For gain weight, the number and percentage of patients who experienced weight gain by ≥+2 kg, ≥+5 kg, or ≥+10 kg over their baseline weight through the DB period, will be summarized by treatment group.

By-patient listings of vital sign measurements will be presented in data listings.

Physical examination results at baseline, V2, V3, V4, V5, V6 and if applicable V7 and V8 will be summarized by treatment group; shifts from Baseline in physical examination findings to each scheduled visit and the worst value through DB period (including scheduled and unscheduled visits) will also be presented.

All physical examination findings will be presented in a data listing.

Liver, and bladder and urinary tract ultrasound results will be summarized descriptively, including the number and percent of patients with normal, abnormal, and clinically significant abnormal results at each applicable study visit by treatment group; shifts from Baseline in Liver, and bladder and urinary tract ultrasound results to each on study visit will also be presented. Data will be presented in a data listing.

4.8.7 Electrocardiogram

In the following calculations, observed data more than 30 days after study treatment discontinuations for patients who discontinued study treatment are handled as missing values (see section 4.2.9), but are included and flagged in any data listings.

12-lead ECG results will be summarized descriptively, including the number and percent of patients with normal, abnormal, and clinically significant abnormal results at Baseline, V4, V6 and if applicable V8; shifts from Baseline in 12-lead ECG results to each scheduled visit and the worst value through DB period (including scheduled and unscheduled visits) will also be presented.

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	Document Date:	10-May-2023

All ECG data for each patient will be provided in data listings.

4.8.8 Concomitant Medications

Any medications that did not stop before the first study treatment date and continued into the DB period will be counted as concomitant medications as well as any medications that start on or after the first study treatment date, whichever its end date.

Concomitant medications will be coded using the WHO Drug dictionary. Results will be summarized by treatment group using Therapeutic Subgroup (ATC level 2) and Chemical Subgroup (ATC level 4). Concomitant medications will be listed in descending order of total frequency within therapeutic and chemical subgroup.

The use of concomitant medications will be included in by-patient data listing.

4.8.9 Concomitant Procedures

Any procedures that did not stop before the first study treatment date and continued into the DB period will be counted as concomitant procedure as well as any procedures that start on or after the first study treatment date, whichever its end date.

Concomitant procedures will be coded using MedDRA and displayed by treatment group in tables and listings using system organ class (SOC) and preferred term (PT). Concomitant procedures will be listed in tables by descending order of total frequency within SOC and PT.

Cytel **Statistical Analysis Plan**

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

CCI

[redacted]

CCI [redacted]

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	**Document Date:**	10-May-2023

CCI

[redacted]

CCI [redacted]

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	Document Date:	10-May-2023

5. CHANGES TO PLANNED ANALYSES

Another secondary endpoint was added (see section 2.12.1): *Proportion of participants achieving sustained improvement in PBC Worst Itch NRS as defined as having responses for all the last three 4-week periods (periods 11, 12 and 13) according to clinically meaningful change; at least 30% reduction; and one-point, two-point or three-point decrease in score from baseline in patients with a baseline PBC Worst Itch NRS score ≥ 4 (binary)*, which is not mentioned in the protocol.

The other secondary endpoint "Proportion of patients with no worsening of pruritus from baseline through week 52 and through week 24 as measured by the PBC Worst Itch NRS" has been re-defined at week 52, i.e. at 4-week period 13 and at week 24, i.e. at 4-week period 6 and will be presented for patients with a baseline PBC Worst Itch NRS score <8.

The other secondary endpoint "Proportion of responders in PBC Worst Itch NRS score according to clinically meaningful change; at least 30% reduction; and one point, two points or three points decrease in score from baseline at week 52 and at week 24 in patients with baseline PBC Worst Itch NRS score ≥ 4" has been re-defined at week 52, i.e. at 4-week period 13 and at week 24, i.e. at 4-week period 6.

6. REFERENCES

Lammers, W.J., K.V. Kowdley, and H.R. van Buuren, Predicting outcome in primary biliary cirrhosis. Ann Hepatol, 2014. 13(4): p. 316-26.

Corpechot C, Abenavoli L, Rabahi N, et al. Biochemical response to ursodeoxycholic acid and long-term prognosis in primary biliary cirrhosis. Hepatology 2008;48:871-7.

Corpechot C, Chazouilleres O, Poupon R. Early primary biliary cirrhosis: biochemical response to treatment and prediction of long-term outcome. J Hepatol 2011;55:1361-7.

Dmitrienko, A, D'Agostino, RB, Huque, MF. (2013). Key multiplicity issues in clinical drug development. Statistics in Medicine. 2013;32:1079-1111.

He X, Chen L, Lei L, Xia HA, Lee MLT (2015) A Simple Method for Estimating Confidence Intervals for Exposure Adjusted Incidence Rate and Its Applications to Clinical Trials. J Biom Biostat 6: 238. doi:10.4172/2155-6180.1000238

Kuiper EM, Hansen BE, de Vries RA, et al. Improved prognosis of patients with primary biliary cirrhosis that have a biochemical response to ursodeoxycholic acid. Gastroenterology 2009;136:1281-7.

Carbone M et al. The UK-PBC risk scores: Derivation and validation of a scoring system for long-term prediction of end-stage liver disease in primary biliary cholangitis. Hepatology. 2016; VOL.63, NO.3.

ICH E9 (R1) addendum on estimands and Sensitivity Analysis in Clinical Trials to the guideline on statistical principles for clinical trials EMA/CHMP/ICH/436221/2017.

Lammers WJ, Hirschfield GM, Corpechot C, et al. on behalf of the Global PBC Study Group. Development and validation of a scoring system to predict outcomes of patients with primary biliary cirrhosis receiving ursodeoxycholic acid. Therapy, Gastroenterology (2015), doi: 10.1053/j.gastro.2015.07.061.

Lammers WJ, van Buuren HR, Hirschfield GM, et al. Levels of alkaline phosphatase and bilirubin are surrogate end points of outcomes of patient with primary biliary cirrhosis: an international follow-up study. Gastroenterology. 2014;147:1338-49.

Lipkovich I, O'Kelly M. Combining Analysis Results from Multiply Imputed Categorical Data. PharmaSUG 2013 - Paper SP03.

Ratitch B, O'Kelly M. Implementation of Pattern-Mixture Models Using Standard SAS/STAT Procedures. PharmaSUG 2011 - Paper SP04.

S. Elman, L.S. Hynan, V. Gabriel, and M.J. Mayo, The 5-D itch scale: a new measure of pruritus, Br J Dermatol. 2010 March ; 162(3): 587–593. doi:10.1111/j.1365-2133.2009.09586.x.

PROMIS®, Scoring manual, A brief guide to scoring the PROMIS® Fatigue instruments. 2021 April.

Mapi Research Trust, Epworth Sleepiness Scale (ESS), Version 1.0, Scaling and Scoring, Version 2.0: August 2017.

A Jacoby, A Rannard, D Buck, N Bhala, J L Newton, O F W James, D E J Jones, Development, validation, and

Sponsor:	GENFIT		
Protocol:	GFT505B-319-1, IND number: 132202		
Document Version No.:	4.0	Document Date:	10-May-2023

evaluation of the PBC-40, a disease specific health related quality of life measure for primary biliary cirrhosis, 16 June 2005, Gut 2005;54:1622–1629. doi: 10.1136/gut.2005.065862.

Lesley A. Inker, M.D., Christopher H. Schmid, Ph.D., Hocine Tighiouart, M.S., John H. Eckfeldt, M.D., Ph.D., Harold I. Feldman, M.D., Tom Greene, Ph.D., John W. Kusek, Ph.D., Jane Manzi, Ph.D., Frederick Van Lente, Ph.D., Yaping Lucy Zhang, M.S., Josef Coresh, M.D., Ph.D., and Andrew S. Levey, M.D. for the CKD-EPI Investigators*, Estimating Glomerular Filtration Rate from Serum Creatinine and Cystatin C, N Engl J Med. 2012 July 5.