[redacted]

Final Analysis
Statistical Analysis Plan (SAP)

ENANTA Pharmaceuticals

Protocol Title:	**A PHASE 2b RANDOMIZED, DOUBLE BLIND, PLACEBO-CONTROLLED, MULTICENTER STUDY EVALUATING SAFETY AND EFFICACY OF EDP-305 IN SUBJECTS WITH LIVER-BIOPSY PROVEN NON-ALCOHOLIC STEATOHEPATITIS (NASH) (ARGON-2)**
Protocol Number:	**EDP 305-102**
Protocol Version, Date	Version 3.0 (Amendment 2.0), 21 March 2021
ICON ID:	3954/0002
Document Version, Date:	FINAL, 14 January 2022

Prepared by:

[redacted]

On behalf of:

ENANTA Pharmaceuticals Inc.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

REASON: I approve this document

792bd426-fd56-475f-9286-659c21213a50

[redacted]

Date
(DD Mmm YYYY)

DocUUID[redacted]

REVISION HISTORY

Version/Date	Version name	Section	Changes implemented
Draft 0.1/ 06-Jul-2020	Initial Draft	N/A	N/A
Draft 0.2/ 07-May-2021	Second Draft	All	Implementing feedback from Enanta review of initial draft.
Final 1.0/ 14-Jan-2022	Final	9	Updated post-study cancellation to identify analyses that will not be produced.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
AE	Adverse Event
ANCOVA	Analysis of Covariance
ASCVD	Atherosclerotic cardiovascular disease
ALT	Alanine aminotransferase
AST	Alanine aminotransferase
ATC	Anatomical Therapeutic Chemical
BLQ	Below limit of quantitation
BMI	Body Mass Index
BP	Blood Pressure
C4	7-alpha-Hydroxy-4-cholesten-3-one
CEC	Clinical Event adjudication Committee
CI	Confidence Interval
CLDQ	Chronic Liver Disease Questionnaire
CM	Concomitant Medication
cm	Centimetre
CNS	Central Nervous System
CRF	Case Report Form
CRO	Clinical Research Organization
CSR	Clinical Study Report
CV	Coefficient of variation
DBL	Database Lock
DBP	Diastolic Blood Pressure
DMC	Data Monitoring Committee
DSMB	Data Safety Monitoring Board
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
ELF	Enhanced Liver Fibrosis
EOS	End of Study
EOT	End of Treatment
FIB-4	Fibrosis-4 score

FSH	Follicle stimulating hormone
HA	Hyaluronic acid
HDL	High density lipoprotein cholesterol
IA	Interim Analysis
ICF	Informed consent form
ICH	International Conference on Harmonisation
IMP	Investigational Medicinal Product
IRT	Interactive Response Technology
ITT	Intent To Treat
IWRS	Interactive Web Response System
kg	Kilogram
LDL	Low density lipoprotein cholesterol
LLN	Lower Limit of Normal
LOCF	Last Observation Carried Forward
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation
MMRM	Mixed Model for Repeated Measurements
MRE	Magnetic resonance elastography
MRI	Magnetic resonance imaging
MRI-PDFF	Magnetic resonance imaging - proton density fat fraction
NAS	NAFLD activity score
NASH	Non Alcoholic Steatohepatitis
OC	Observed Cases
PCS	Potentially Clinically Significant
PD	Pharmacodynamics
PI	Principal Investigator
PIIINP	Procollagen III amino terminal peptide
PK	Pharmacokinetics
PR	electrocardiographic interval occurring between the onset of the P wave and the QRS complex, representing time for atrial and ventricular depolarization, respectively
PP	Per Protocol
PRO-C3	Type 3 procollagen
PT	Preferred Term

Q1	Lower Quartile
Q3	Upper Quartile
QLS	Quality of Life Scale
QoL	Quality of Life
QRS	electrocardiographic deflection between the beginning of the Q wave and termination of the S wave, representing the time for ventricular depolarization
QT	electrocardiographic interval between the beginning of the Q wave and termination of the T wave, representing the time for both ventricular depolarization and repolarization to occur
QTcF	QT interval corrected by Fridericia's formula
RR	interval between successive heart beats using the R-wave peaks
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SBP	Systolic Blood Pressure
SD	Standard Deviation
SI	Standard International
SOC	System Organ Class
SOP	Standard Operating Procedure
SSP	Study Specific Procedure
T2DM	Type 2 diabetes mellitus
TC	Total cholesterol
TEAE	Treatment Emergent Adverse Event
TFLs	Tables, Figures and Listings
TG	Triglyceride
TIMP-1	Tissue inhibitor of metalloproteinase 1
ULN	Upper Limit of Normal
VAS	Visual Analog Score
WHODDE	World Health Organization Drug Dictionary Enhanced
WOCF	Worst Observation Carried Forward

1 INTRODUCTION

The Statistical Analysis Plan (SAP) describes the statistical methods to be used during the reporting and analyses of data collected under Enanta Pharmaceuticals, Inc. Protocol EDP 305-102 entitled "*A Phase 2b Randomized, Double Blind, Placebo-Controlled, Multicenter Study Evaluating Safety and Efficacy of EDP-305 in Subjects with Liver Biopsy Proven NonAlcoholic Steatohepatitis (NASH)*".

This SAP should be read in conjunction with the study protocol and case report form (CRF). This version of the plan has been developed using the protocol version 3.0 (Amendment 2.0) dated 21March2021. Any further changes to the protocol may necessitate updates to the SAP.

Any deviations from this SAP will be described and justified in the Clinical Study Report (CSR).

The preparation of this SAP has been based on International Conference on Harmonization (ICH) E9 guidelines. The table of contents and shells for the tables, figures and listings (TFLs) will be produced in a separate document. All data analyses and generation of TFLs will be performed using SAS 9.4® or higher.

2 STUDY OBJECTIVES

2.1 Primary objective

To evaluate the effect of EDP-305 compared to placebo on liver histology in non-cirrhotic NASH subjects with stage 2 or 3 fibrosis.

2.2 Secondary objectives

- To evaluate the effect of EDP-305 on liver histology by assessing
  - Improvement of fibrosis by at least 1 stage and/or resolution of NASH, without worsening of either
  - No worsening of fibrosis and no worsening of NASH
  - Resolution of fibrosis
  - Improvement in each histological feature of NASH by at least 1 point
  - Improvement of fibrosis by ≥ 2 stages
  - Improvement in NAS by at least 2 points with no worsening of fibrosis
  - Improvement of fibrosis and resolution of NASH as a composite endpoint and as defined by both endpoints being met in the same subject
  - Resolution of NASH and no worsening of liver fibrosis
  - Histological progression to cirrhosis based on the overall assessment made
- To evaluate the safety of EDP-305
- To evaluate the effect of EDP-305 on pruritus
- To evaluate the effect of EDP-305 on hepatic steatosis
- To evaluate the effect of EDP-305 on liver stiffness
- To evaluate the effect of EDP-305 on lipid profile
- To evaluate the pharmacokinetics (PK) of EDP-305 and its metabolites in plasma

[redacted]

3 STUDY DESIGN

3.1 General study design

This is a phase 2b randomized, double blind, placebo-controlled multicenter study evaluating the safety and efficacy of two doses of EDP-305 compared to placebo in subjects with liver biopsy proven NASH.

The duration of the study will be approximately up to 84 weeks. The study will consist of a Screening, Treatment, and Safety Follow-up Periods as below.

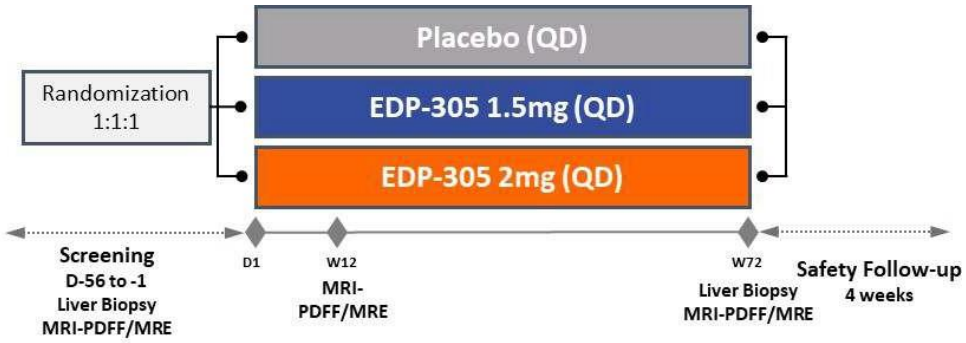
Study Period	Duration
Screening	Up to 8 weeks (56 Days)
Treatment	72 weeks
Safety Follow-up Period	4 weeks
Total approximate duration of participation	up to 84 weeks

- Screening period: includes the screening visit and may occur over a time period of 56 days (8 weeks) prior to the first dose of study drug.
- Treatment period: begins with the first dose of study drug on Day 1 and concludes with the end of treatment (EOT) visit. Subjects who complete dosing will have their EOT visit during Week 72.
- Safety Follow-up period: commences following the last dose of study drug and concludes at the end of study (EOS) visit. Subjects who complete the safety follow-up period will have their EOS visit during Week 76.

3.2 Dose and Treatment Schedule

Subjects will be randomized 1:1:1 to receive one of two oral doses of EDP 305 1.5 mg, EDP-305 2 mg or matching placebo. Every subject will receive a single daily dose of blinded study drug for a total of 72 weeks. An overview of the study design is shown in Figure 1. Study visits and assessments are detailed in the Table 1.

Figure 1: Study Design



3.3 Randomization and blinding

3.3.1 Randomization

Subjects will be randomized to study treatment using an interactive web response system (IWRS). Subjects will be randomized to the treatment groups shown below:

- Treatment Group 1 (N=112); EDP-305 1.5 mg orally for 72 weeks
- Treatment Group 2 (N=112); EDP-305 2 mg orally for 72 weeks
- Treatment Group 3 (N=112); Placebo orally for 72 weeks

In addition, subjects will be stratified based on 1) use of Vitamin E and/or pioglitazone and 2) type 2 diabetes mellitus (T2DM) status.

The randomization code will be produced by Enanta (or designee). The Enanta unblinded biostatistician or designee will review and approve the final randomization list.

During the screening period, subjects will be identified by a unique screening number assigned by the clinical site. Subjects who have completed screening assessments and are eligible for participation in the study will be randomized before the first dose of study drug (Day −1 or Day 1) and assigned a unique subject number which will be used to identify the subject throughout the study.

3.3.2 Blinding

The study will be double-blinded meaning the subjects, Sponsor, PIs, and site staff will be blinded to treatment assignment until the completion of the study.

All study personnel will be blinded to treatment assignment except for the following individuals:

- Unblinded Enanta/Clinical Research Organization (CRO) statistician for purpose of generating and monitoring the randomization list
- Unblinded Drug Supply Chain personnel for the purpose of monitoring drug supplies

- Enanta/[redacted] and Regulatory Affairs representatives when required to satisfy regulatory reporting requirements
- Bioanalytical Laboratory for the purpose of measuring drug concentrations
- Data Safety Monitoring Board (DSMB) members or DSMB meeting attendees (including Enanta representatives not associated with the day-to-day conduct of the study) for the purposes of unblinded data review as outlined in the DSMB charter

Refer Section 6.1 of the Study Protocol for Blinding of Study and Pharmacokinetic samples.

Unblinding of individual subject treatment by the PI should be limited to medical emergencies or urgent clinical situations in which knowledge of the subject's study treatment is necessary for clinical management. In such cases, where possible, the the PI should attempt to contact the study Medical Monitor to discuss the need for unblinding. In situations in which the PI has attempted and failed to contact the Medical Monitor, and/or the urgency of the case required immediate action, PIs should use their best judgment, based on the nature and urgency of the clinical situation, and proceed with unblinding. Once a subject's treatment assignment has been unblinded for a medical emergency or urgent clinical situation, the Medical Monitor should be notified within 24 hours of unblinding of the treatment. Information relating to unblinding (e.g., the reason, date) should be clearly recorded in the subject's study file.

3.4 Study treatments and assessments

Subjects who meet all criteria for enrollment will be randomized to blinded treatment on Day 1 in a 1:1:1 ratio to EDP-305 2 mg, EDP-305 1.5 mg, or placebo. Assignment to treatment groups will be determined by a computer-generated random sequence using an IWRS. The IWRS will be used to assign double-blind investigational product to each subject. To achieve between-group comparability the randomization will be stratified by diabetes status and use of Vitamin E and/or pioglitazone.

A detailed description of procedures and assessments to be conducted during this study is summarized in the scheduled of study assessments in table below.

Table 1: Schedule of Study Assessments

Study Event	Screening[1]	Study Assessments per Planned Study Day/Week										Phone Contact[2]	EOT	EOS
Visit Day/Week	D-56 to D-1	D1[3]	D7[4]	W2	W4	W8	W12	Wks 16, 20, 24, 28, 32	W36	W40 & W44	W48, W56, W64 (Every 8 weeks)	W52, W60, W68	W72[5]	W76
Visit Windows			±1d	±2d	±2d	±2d	±3d	±3d	±3d	±3d	±3d		±3d	±3d
ICF[6]; Demography; Medical History	x													
Inclusion/Exclusion	x													
Liver Biopsy (with centralized reading)	x[7]												x	
Elastrography: Liver Stiffness Assessment (MRE)	x						x						x	
NASH FibroSure®	x	x					x						x	
APRI, FIB-4, and NFS		x					x						x	
FSH[8]; HIV, HCV, and HBV	x													
Pregnancy Test[9]	x	x			x	x	x	X	x	x	x		x	x
At Home Pregnancy Tests (Phone Contact) [10]												x		
Height, Weight and BMI[11]	x	x					x		x				x	x
Physical Exam[12]	x	x		x	x	x	x	X	x	x	x		x	x
Vital Signs[13]	x	x		x	x	x	x	x	x	x	x		x	x
Body Temperature	x	x												x
ECG	x	x					x	W24 only	x		W48 & W64 only		x	
Safety Lab. Tests[14]	x	x		x	x	x	x	x	x	x	x		x	x
Pruritus: Visual Analogue Score, 5D-itch scales[15]		x	x	x	x	x	x	x	x	x	x		x	x
QoL Scales[16]		x					x	W24 only	x		W48 only		x	x
Cardiovascular Score (ASCVD)		x											x	
PT/PTT and INR	x	x		x	x	x	x	x	x	x	x		x	x
CV Markers[17]		x		x	x	x	x	x	x	x	x		x	x
MRI-PDFF	x						x						x	
Inflammatory Markers[18, 19]		x					x						x	
ELF Panel, PRO C3 [17]		x					x						x	

FGF-19, C4, BA[20]		x		x	x	x	x	x	x	x	x		x	
PK samples[21]		x		x	x	x	x	x	x	x	x		x	
CK-18 and GLP-1[17]		x					x	W24 only	x		W48 only		x	
Study Drug Dosing[22]		Daily Dosing												
Drug Accountability		x	x	x	x	x	x	x	x	x	x	x	x	
Concomitant Medication	x	x	x	x	x	x	x	x	x	x	x	x	x	x
AE/SAE	x	x	x	x	x	x	x	x	x	x	X	x	x	x
Exploratory research samples[17]		x					x	W24 only	x		W48 only		x	

[1] Screening assessments should be conducted within 56 days prior to the first dose of study drug (ie, Study Days -56 to -1)

[2] At Weeks 52, 60 and 68, the site will contact the subject by phone to verify results of the home pregnancy test (for female subjects of child-bearing potential) and assess drug accountability, concomitant medications and AE/SAEs.

[3] On Day 1, all samples are to be collected predose; two additional PK and PD samples will be collected postdose as noted in protocol Sections 8.4.4 and 8.4.5.

[4] Day 7 is a phone contact, not an in-clinic visit.

[5] Subjects should take their last dose of study drug at the Week 72 visit. Subjects who return for their EOT Visit after Week 72, should stop dosing during Week 72 as instructed by the site. Subjects who discontinue the study early should complete the EOT procedures within one week following discontinuation of study drug; For subjects who had the MRI and MRE conducted at the Week 12 visit but discontinue treatment prior to Week 36, no additional MRI and MRE will be conducted. For subjects who discontinue treatment at Week 36 or later, the MRI and MRE should be conducted within 2 weeks after discontinuing treatment, and only one PK sample will be obtained.

[6] Informed consent must be obtained prior to conducting any study-specific procedures or assessments.

[7] The biopsy may be obtained either 1) during the Screening window or 2) within 6 months prior to the Screening visit. If done during the study screening window, the biopsy should be performed once all other I/E criteria have been met.

[8] For post-menopausal women only.

[9] Serum pregnancy test at Screening and Baseline and urine pregnancy testing at Baseline and all other visits for all females. If the urine pregnancy test is positive, a serum pregnancy test should be obtained as soon as possible. After Baseline, pregnancy tests do not need to be performed on females who are confirmed to be of non-childbearing potential. A serum pregnancy test must also be completed at the EOS visit for females who are of child-bearing potential.

[10] Urine pregnancy tests will be provided to females of child-bearing potential so that they can perform the test at home on Weeks 52, 60 and 68.

[11] Height to be assessed at Screening only.

[12] Full physical exam (PE) at Screening, Week 12, and EOS Visit; subsequent PE should be targeted to review new signs and symptoms.

[13] Vital Signs include heart rate, respiratory rate, blood pressure, and will be measured predose.

[14] Safety laboratory tests include chemistry (including liver function tests), hematology, and urinalysis and should be collected predose at all visits; See Table 3 from protocol for details. eGFR will be calculated at all visits based on the MDRD formula. HbA1c will be obtained at Screening, Baseline, week 12, and Week 72.

[15] Pruritus scales (up to 2 scales) will be utilized to measure pruritus for every subject regardless of whether or not a subject experiences pruritus. Please refer to protocol Section 8.3.5 and additional details included in the study binder.

[16] QoL includes SF-36 and CLDQ-NASH. In addition, Dermatology Life Quality Index (DLQI) will be completed at each visit by subjects who reported pruritus until resolution. Additional details can be found in the study binder.

[17] Lipids and CV risk markers to be collected are detailed in Table 3 from protocol.

[18] Markers of inflammation include fibrinogen, CRP, IL-6, IL-1β, TNF-α, TNF-β, alpha2 macroglobulin and haptoglobin (See Table 3 from protocol).
[19] One sample will be collected from all subjects predose.
[20] Samples should be collected before the subject takes the daily dose of study drug. On Day 1 and Weeks 12, 36, and 72, samples should be collected predose and two samples postdose; the first postdose sample will be collected 1 to 3 hours after dosing and the second collected at least 1 hour later. At all other visits, samples will be collected only predose. If the subject took drug prior to their clinic visit, only one postdose sample should be collected.
[21] PK predose samples should be collected before the daily dose of study drug. On Day 1 and Weeks 12, 36, and 72, one PK sample should be collected predose and two samples postdose; the first postdose sample should be collected 1 to 3 hours postdose and the second postdose sample at least 1 hour later. At all other visits, PK samples will be collected predose. If the subject took drug prior to their clinic visit, only one postdose sample should be collected. For subjects who discontinue treatment early, collect one postdose sample at the EOT visit. For subjects with persistent transaminase or ALP elevations and evidence of liver injury and who remain on study drug, the one additional PK sample will be collected at each visit where safety labs are obtained.
[22] See protocol Section 5.7 for details.

4 STUDY ENDPOINTS

4.1 Primary efficacy endpoints

There are two primary efficacy endpoints for this study: (1) the proportion of subjects who achieve ≥1 stage improvement in fibrosis without worsening of steatohepatitis as determined by liver biopsy at 72 weeks, (2) the proportion of subjects with resolution of steatohepatitis and no worsening of liver fibrosis as determined by liver biopsy at 72 weeks *[Note: No worsening of steatohepatitis is defined as no increase in NAS for ballooning, inflammation, or steatosis. Resolution of steatohepatitis is defined as absent fatty liver disease or isolated or simple steatosis without steatohepatitis and a NAS score of 0–1 for inflammation, 0 for ballooning, and any value for steatosis]* (Agopian, et.al.)

4.2 Secondary efficacy endpoints

The secondary endpoints for this study are:

- Proportion of subjects with improvement of fibrosis by at least 1 stage and/or resolution of NASH without worsening of either as determined by liver biopsy at week 72
- Proportion of subjects with no worsening of fibrosis combined with no worsening of NASH as determined by liver biopsy at week 72
- Proportion of subjects with resolution of fibrosis as determined by liver biopsy at week 72
- Proportion of subjects with improvement in each histologic feature of NASH, by at least 1 point as determined by liver biopsy at week 72
- Proportion of subjects with improvement of fibrosis by ≥ 2 stages by liver biopsy at week 72
- Proportion of subjects with improvement in NAS by at least 2 points with no worsening of fibrosis as determined by liver biopsy at week 72
- Proportion of subjects with improvement of fibrosis and resolution of NASH as a composite endpoint as defined by both endpoints being met in the same subject
- Proportion of subjects with resolution of NASH and no worsening of liver fibrosis
- Proportion of subjects with histological progression to cirrhosis as determined by liver biopsy at week 72
- Frequency of adverse events (AEs), serious adverse events (SAEs), and AEs leading to discontinuation through week 72 and 4-week follow-up period
- Change from baseline in 5D-itch scale and Visual Analog Score (VAS) through week 72
- Change from baseline in percentage of fat in the liver as assessed by magnetic resonance imaging proton density fat fraction (MRI PDFF) at week 12 and week 72

- Change from baseline in liver stiffness as assessed by MRE at week 12 and week 72
- Change from baseline in triglyceride (TG), total cholesterol (TC), high density lipoprotein (HDL) cholesterol, low density lipoprotein (LDL) cholesterol and adiponectin through week 72
- Pharmacokinetic concentrations of EDP-305 (and metabolites)

[redacted]

4.4 Safety endpoints

- Treatment compliance
- Adverse events
- Clinical laboratory data
- Electrocardiogram data
- Concomitant medications
- Vital Signs
- Physical examinations

[redacted]

6 ANALYSIS POPULATIONS

6.1 Safety population (SAF)

All subjects who receive at least one dose of study drug. Subjects will be included in the treatment group that corresponds to the first study drug received during the study.

6.2 Intention-To-Treat (ITT) population

ITT subjects will be considered those randomized to treatment. The ITT subjects will be analyzed according to the treatment to which they were randomized. In the event the safety population is the same as ITT, only the ITT will be reported. This will be the primary efficacy population.

6.3 Per-Protocol population (PP)

PP population includes all subjects in the ITT population that do not have major protocol deviations which could unduly influence the efficacy analysis. This population will be reviewed and finalized prior to locking the database and unblinding the study.

The per protocol population include patients who:

- Satisfying all inclusion and exclusion criteria,
- Receiving the full treatment regimen (i.e. all doses received and at least 80% of protocol-specified total dose taken for the entire treatment phase),
- Did not take any prohibited medications within *X* weeks of the primary study time point (i.e., Week 72), and
- Have a value for at least one of the primary endpoints.

Subjects will be included in the analysis according to the treatment to which they were randomized. The list of subjects in the PP population will be finalized prior to unblinding.

6.4 Pharmacokinetic (PK) Population:

All subjects receiving active study drug and having any measurable plasma concentration of study drug at any time point.

Subjects will be included in the analysis according to the study drug received during the study.

6.5 Protocol deviations/violations and exclusions from analysis sets

Protocol deviations will be identified following the ICON SOPs, and after review and validation of the Enanta study team prior to locking the database and unblinding the treatment codes. If necessary, a meeting will be held to discuss and adjudicate any deviations that need classification (major vs. minor). Deviations will be identified using data listings provided by ICON Biometrics and from the deviation log maintained by ICON clinical.

The listing of protocol deviations will be reviewed by Enanta team with focus from Biometrics/Medical in order to define the Per Protocol population. Subjects with a protocol violation or a major protocol deviation will be thus identified for exclusion from the per protocol population.

7 STATISTICAL CONSIDERATIONS AND ANALYSIS

7.1 Derived variables

The below table provides the list of derived variables for demographic and baseline characteristics, various duration derivations, drug compliance, baseline derivations and other important derivations applicable for this study.

Table 2: Derived variables.

Variables	Formula
Demographic and Baseline characteristics	
Age at informed consent (in years)	(Date of informed consent – Date of birth + 1)/ 365.25)
Body mass index (BMI) (kg/m2)	Weight (kg) / [Height (m)]^2
Derivation of Duration	
Study day at any visit	Date of interest – Date of first dose of study drug. One day is added if this difference is ≥ 0
Extent of exposure (Days)	Date of last randomized study drug intake – Date of first randomized study drug intake + 1
Extent of exposure (Weeks)	Extent of exposure (days)/7
Drug Compliance	
Compliance	100 × ((Amount of drug taken) / Exposure period)
Baseline Derivations	
Baseline	Last non-missing values collected prior to the first dose of study drug on day 1
Screening	The value obtained at screening.
Change from baseline	Post baseline value – Baseline
Percentage change from baseline	[(Change from baseline) / Baseline] * 100
Reference baseline	Mean of screening and day 1 value
Other Derivation	
Improvement in liver fibrosis	≥1 point decrement in fibrosis score compared to baseline
No worsening in liver fibrosis	No increment in fibrosis score compared to baseline
Resolution of liver fibrosis	If fibrosis score is 0
Improvement in NAS score	≥1 point decrement in NAS score (can occur in any of the ballooning, inflammation, or steatosis categories)

No worsening in NAS score	No increase in NAS score for ballooning, inflammation, and steatosis all
Resolution of steatohepatitis	Defined as total absence of ballooning [score = 0], absent or mild inflammation [score 0-1], steatosis can be present [score 0-3]
Platelet Ratio Index (APRI)	APRI = ([AST (IU/L) / AST ULN (IU/L)] / [Platelets(10^9/L)]) × 100 ULN = upper limit of normal Online calculator can be found at: http://www hepatitisc.uw.edu/page/clinical-calculators/apri
Fibrosis-4 score (FIB-4)	FIB-4 = [Age (years) × AST (IU/L)] / [Platelets (109/L) × (√ ALT (IU/L)) FIB-4 online calculator reference: http://www hepatitisc.uw.edu/page/clinical-calculators/fib-4
NAFLD fibrosis score (NFS)	NFS = -1.675 + 0.037 × Age (years) + 0.094 × BMI (kg/m2) + 1.13 × IFG/Diabetes (yes = 1, no = 0) + 0.99 × AST/ALT ratio – 0.013 × Platelets (109/L) – 0.66 × Albumin (g/dL) NFS online calculator reference: https://www.mdcalc.com/nafld-non-alcoholic-fatty-liver-disease-fibrosis-score
Total NAS score	Sum of scores for steatosis, lobular inflammation, and ballooning
Histological progression to cirrhosis	If Fibrosis score = 4
5D-itch scale	Refer Appendix B for derivation
Visual Analog Score (VAS)	Refer Appendix B for derivation
SF-36	Refer Appendix B for derivation
CLDQ-NASH	Refer Appendix B for derivation
ASCVD Score	Refer Appendix B for derivation
QTc	QT/(RR interval)$^{1/3}$, in seconds (The QT/QTc interval will be corrected based on Fridericia's correction)

7.2 Handling of missing data and outliers

7.2.1 Missing data analysis methods

For the primary analysis of the primary endpoints, no imputation will be performed (i.e., only subjects with endpoint values will be utilized). A sensitivity analysis will be conducted where subjects with missing primary endpoint data will be imputed as non-responders.

7.2.2 Handling of missing or incomplete dates

Imputation rules for missing or partial AE start date are defined below:

If only Day of AE start date is missing:

If the AE start year and month are the same as that for the first dose date, then:

- If the full (or partial) AE end date is NOT before the first dose date or AE end date is missing, then impute the AE start day as the day of first dose date;
- Otherwise, impute the AE start day as 1.

If Day and Month of AE start date are missing:

If AE start year = first dose year, then:

- If the full (or partial) AE end date is NOT before the first dose date or AE end date is missing, then impute the AE start month and day as the month and day of first dose date;
- Otherwise, impute the AE start month as January and the day as 1.

If Year of AE start date is missing:

If the year of AE start is missing or AE start date is completely missing, then query site and leave as missing.

For missing and partial adverse event end dates, imputation will be performed as follows:

If only the day of the month is missing, the last day of the month will be used to replace the missing day. If the day and month are missing or a date is completely missing, it will be considered as missing.

Imputation rules for missing or partial medication start/stop dates are defined below:

If only Day of CM start date is missing:

If the CM start year and month are the same as that for the first dose date, then:

- If the full (or partial) CM end date is NOT before the first dose date or CM end date is missing, then impute the CM start day as the day of first dose date;
- Otherwise, impute the CM start day as 1.

If Day and Month of CM start date are missing:

If CM start year = first dose year, then:

- If the full (or partial) CM end date is NOT before the first dose date or CM end date is missing, then impute the CM start month and day as the month and day of first dose date
- Otherwise, impute the CM start month as January and the day as 1.

If Year of CM start date is missing:

- If the year of CM start is missing or CM start date is completely missing, then query site and leave as missing.

8 STATISTICAL METHODS

8.1 General statistical conventions

All statistical procedures will be completed using SAS version 9.4 or higher.

Unless otherwise stated, all statistical testing will be two-sided and will be performed using a significance (alpha) level of 0.05. Two-sided 95 % confidence intervals (CI) will be provided when relevant. Any p-value below 0.05 will be displayed to 4 decimal places.

Any p-value less than 0.0001 will be displayed as <0.0001. Any p-value between 0.05 (inclusive) and 0.10 will be displayed to 3 decimal places and any p-value greater than 0.10 will be displayed to two decimal places. Any p-value greater than 0.9999 will be displayed as >0.99.

All quantitative endpoints will be summarized using an 8-number summary (n, mean, standard deviation, median, 25th percentile, 75th percentile, minimum and maximum values). The mean and median and the percentiles will be rounded to one decimal place beyond the precision of the values being summarized, the standard deviation will be rounded to 2 additional decimal places beyond this precision and the minimum and maximum values will be displayed in the same precision. All qualitative endpoints will be summarized by the number of subjects meeting the endpoint and the percentage of subjects out of the appropriate population. The denominator will be displayed when needed and the percentage will be rounded to one decimal place

All subject data, including those derived, will be presented in individual subject data listings. Unless otherwise stated, unscheduled visit results will be included in date/time chronological order, within subject listings only. All listings will be sorted by investigational site, patient number, date/time, and visit. The treatment group as well as patient's sex and age will be stated on each listing. Unless otherwise stated, data listings will be based on all subjects randomized.

8.2 Subject disposition

All subjects who provided informed consent will be included in a summary of subject accountability. The number and percentage of subjects screened, randomized, randomized and treated, randomized and not treated, as well as the number and percentage of subjects in each analysis population (safety, ITT, PP, and PK) will be summarized. The denominator for the calculation of percentages will be from the number of subjects randomized.

The following categories will also be summarized for subject disposition by treatment group and overall:

- Completed study drug per protocol
- Discontinued study drug early and the primary reason for study drug discontinuation
- Completed the study (i.e., completed safety follow-up at Week 76)
- Discontinued from the study early and the main reason for discontinuation

8.3 Protocol deviations

All protocol deviations identified will be summarized by treatment group and overall, all protocol deviations will be listed.

Summaries of major protocol deviations (counts & percentages) will be provided by deviation categories & subcategories for all subjects that were randomized.

8.4 Demographics and baseline characteristics

No statistical testing will be performed for the comparison between treatment groups on demographics and baseline characteristics.

A by-subject listing will be provided.

8.4.1 Demographics

The following subject demographics will be summarized in the table by randomized treatment group for all subjects in the safety population: age, gender, ethnicity, race, height, weight and BMI at baseline.

8.4.2 Baseline and disease characteristics

T2DM status, use of Vitamin E and/or pioglitazone, fibrosis score, NAS score, individual NAS components (steatosis, lobular inflammation, ballooning) at baseline will be summarized for the safety population. The categorical baseline characteristics such as ECG and selected lab parameters (i.e., ALT, AST, Total Bilirubin, LDL, HDL, C4, and FGF) will be summarized using frequency counts for the safety population.

8.4.3 Medical history

A summary of medical and surgical history will be presented by system organ class (SOC) and preferred term (PT) using Medical Dictionary for Regulatory Affairs® (MedDRA) Version 23.0 or higher for the safety population.

A by-subject listing will be provided.

8.4.4 Prior and concomitant medications/ procedures

Medications used in this study will be coded by using the World Health Organization Drug Dictionary Enhanced (WHODDE) version "WHO Drug Global B3 Mar2021".

Prior medications/ procedures: are defined as those medications/ procedures with a start date prior to the first dose of study drug.

Concomitant medications/ procedures: are defined as those medications/ procedures with a start date on or after the first dose of study drug until the last dose. A medication/ procedure which started prior to dosing and continued after dosing will also be considered as concomitant medication/ procedure.

Prior and Concomitant medications will be summarized descriptively using frequency tables by anatomical therapeutic chemical (ATC) class and preferred name by treatment group on the safety population. Concomitant diabetes medications will be summarized using the same format.

Prior and Concomitant procedures will be summarized descriptively using frequency tables by treatment group on the safety population.

A by-subject listing will be provided for prior and concomitant medications.

Details for imputing missing or partial start and/or stop dates of medication are described in Section 7.2.2.

8.5 Extent of exposure

8.5.1 Treatment duration

Duration of study drug (in days) will be calculated as: last dose date – first dose date + 1 day, regardless of study drug interruption.

Study drug exposure will be summarised by treatment group on the safety population using descriptive statistics.

8.5.2 Treatment compliance

Study drug compliance will be based on the amount of drug taken. The derivation for the treatment compliance is provided in Table 2. The maximum percentage of drug taken will be 100%.

Study drug compliance will be summarized by treatment group using 8-number summary. They will also be summarized in categories "<80%, 80% - <90%, 90% - <100% and "≥100% compliant" using frequency tables.

Study drug compliance summaries will be based on the safety population.

A by-subject listing of randomization schedule and study drug dispensed with LOT number will be provided.

8.6 Efficacy analyses

This section addresses separately the analyses to be conducted on the primary, secondary, and exploratory efficacy variables. All the efficacy analyses will be performed using ITT population as primary analysis.

8.6.1 Analysis methods

Logistic Regression Model for proportion of responders

Endpoints looking at a proportion of responders will be analyzed by logistic regression. The SAS procedure PROC GENMOD will be used. The preferred model will include the fixed categorical effects of treatment group, diabetes status and use of Vitamin E and/or pioglitazone. Baseline NAS score will be included as covariate. The model will be used to derive odds ratio, confidence interval and p value. A sample of SAS code for logistic regression is provided in Appendix A.

Fisher exact test will be used to test the treatment difference in case the model does not converge. A sample of SAS code for Fisher exact test is provided in Appendix A.

Analysis of Covariance (ANCOVA) Model for Change from baseline

In summaries of efficacy endpoints examining change from baseline at week 12 and week 72, ANCOVA of the differences between post-baseline and baseline measurements will be performed, with treatment group, baseline diabetes status and use of Vitamin E and/or pioglitazone as fixed effect and the baseline measurement of the parameter of interest along with baseline NAS score as covariate. A sample of SAS code for ANCOVA model is provided in Appendix A.

The model will provide the least-squares means (using type III sum-of-squares), standard errors, p-value and 2-sided 95% confidence interval for treatment differences.

Mixed Model Repeated Measures (MMRM) analysis

A MMRM analysis will be performed for secondary efficacy endpoints that are continuous in nature and assessed over multiple study visits. The SAS procedure PROC MIXED will be used. The preferred model will include the fixed categorical effects of treatment group, visit, treatment-by-visit interaction as well as the continuous fixed covariates of baseline measurement for the parameter of interest and baseline NAS score. An unstructured variance-covariance structure will be considered to model the within-patient error. If the unstructured covariance matrix results in a lack of convergence, the heterogeneous Toeplitz covariance structure should be used. The Kenward-Roger method will be used to estimate the denominator degrees of freedom. A sample of SAS code for MMRM is provided in Appendix A.

The model will provide least-squares means (using type III sum-of-squares), standard errors, p-value and 2-sided 95% confidence interval for treatment differences for each visit.

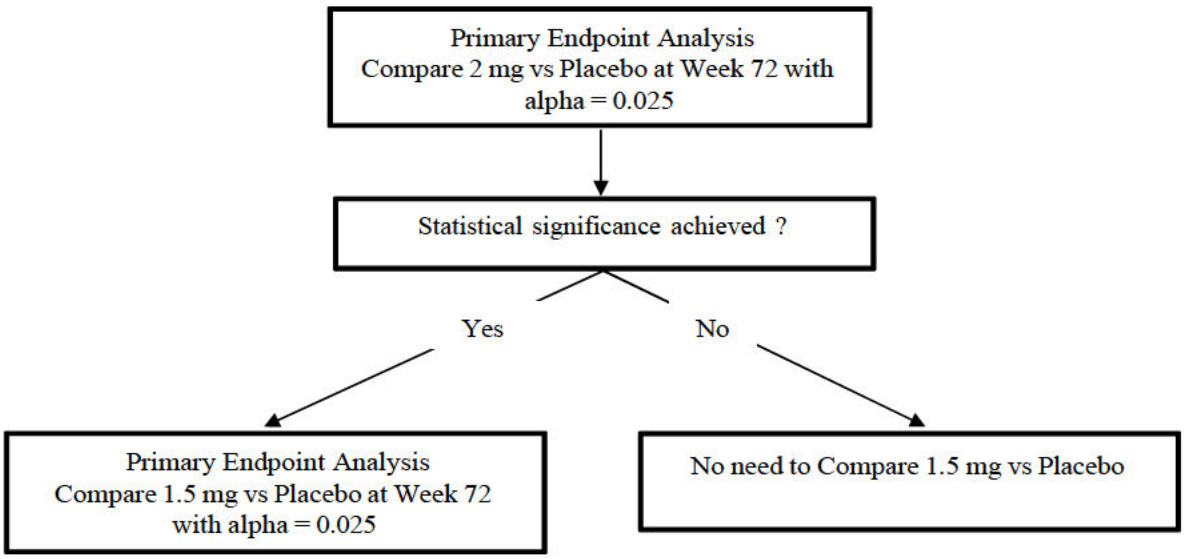
8.6.1.1 Multiplicity

Multiplicity adjustment will be made for primary analysis in this study.

Strongest control of type I error at 0.05 will include an equal Bonferroni split alpha (0.025) for each of the primary endpoints. Multiple active dose groups will be compared to placebo (high dose versus placebo, low dose versus placebo). The high dose group for each will be analyzed first for each endpoint. If statistical significance at the 0.025 level is achieved the low dose group will be tested. If statistical significance is not achieved for the high dose, the low dose group will not be tested.

The Flow Chart of Testing Strategy is presented in Figure 2. This flow chart is for each of the two primary endpoints.

Figure 2: Flow Chart of Testing Strategy



8.6.1.2 Treatment by center interaction analysis (multi-center study)

No analysis will be made to assess the treatment-by-center interaction.

8.6.2 Analysis of primary efficacy endpoint(s)

The primary efficacy analysis is a set of two primary endpoints. The primary endpoints comprise of proportion of subjects who achieve:

- ≥1 stage improvement in liver fibrosis without worsening of steatohepatitis
- Resolution of steatohepatitis without worsening of liver fibrosis

as determined by liver biopsy at week 72. Derivation for steatohepatitis and fibrosis conditions (like improvement, worsening and resolution) are provided in Table 2.

A responder will be defined for each primary endpoint separately. Any subject achieving the primary endpoint will be a responder for that primary endpoint. Subjects will not be categorized as responder or non-responder in case of missing response (score).

The objective related to the primary endpoint is to evaluate the effect of EDP-305 compared to placebo with respect to proportion of responders defined as above.

The hypothesis to test EDP-305 2 mg and placebo is formulated as:

- H_{01}: OR =1 v/s H_{11}: OR ≠ 1

Where OR is the odds ratio at week 72 for EDP-305 2mg and placebo treatment groups respectively.

The hypothesis to test EDP-305 1.5 mg and placebo is formulated as:

- H_{02}: OR =1 v/s H_{12}: OR ≠ 1

Where OR is the odds ratio at week 72 for EDP-305 1.5mg and placebo treatment groups respectively.

The proportion of responders for each endpoint will be compared using a logistic regression model with treatment group, diabetes status and use of Vitamin E and/or pioglitazone as fixed effects and baseline NAS score as covariate. Treatment groups will be compared for each endpoint at a 0.025 significance level. The odds ratio, 97.5% confidence interval and the p-value will be presented. In the event that the logistic regression model is unable to converge, a Cochran-Mantel-Haenszel test will be performed using the stratification factors of diabetes status and Vitamin E/pioglitazone use status. The frequency counts for responders and non-responders will also be presented in the table. The primary efficacy analyses will be performed using the ITT population. As an additional analysis, the primary endpoint will also be analyzed using PP population. Multiplicity control for this testing is noted in Section 8.6.1.1.

A by-subject listing will be provided for the primary efficacy endpoints.

8.6.2.1 Sensitivity Analysis

As a sensitivity analysis, the methodology utilized for the primary endpoints will be performed on all ITT subjects where those missing primary endpoint data will be imputed as a non-responder.

8.6.3 Analysis of secondary efficacy endpoint(s)

All analysis of secondary efficacy endpoints will be performed on the ITT population. Secondary endpoint related to PK data will be analyzed using subjects in the PK population. Secondary endpoints will be analyzed for treatment group differences.

8.6.3.1 Categorical Endpoints (Responder at Week 72)

Secondary endpoints with categorical values are defined as below:

Proportion of subjects with

- improvement of fibrosis by at least 1 stage and/or resolution of NASH without worsening of either as determined by liver biopsy at week 72
- no worsening of fibrosis combined with no worsening of NASH as determined by liver biopsy at week 72
- resolution of fibrosis as determined by liver biopsy at week 72
- improvement in each histologic feature of NASH, by at least 1 point as determined by liver biopsy at week 72
- improvement of fibrosis by ≥ 2 stages by liver biopsy at week 72
- improvement in NAS by at least 2 points with no worsening of fibrosis as determined by liver biopsy at week 72
- improvement of fibrosis and resolution of NASH as a composite endpoint as defined by both endpoints being met in the same subject
- resolution of NASH and no worsening of liver fibrosis
- histological progression to cirrhosis as determined by liver biopsy at week 72

Any subject who satisfies the criteria for the secondary endpoint will be a responder for that endpoint.

The proportion of responders for each endpoint will be compared using a logistic regression model with treatment group, diabetes status and use of Vitamin E and/or pioglitazone as fixed effects and baseline NAS score as covariate at week 72. Treatment groups will be compared for each endpoint at a 0.05 significance level. Odds ratio, 95% confidence interval and p value will be reported in the table.

A Cochran-Mantel-Haenszel test stratifying by diabetes status and Vitamin E/pioglitazone usage will be utilized to test the treatment group differences in case of lack of convergence. The frequency counts for responders and non-responders along with the p-value will be presented in the table.

A by-subject listing will be provided for observed and derived variables.

8.6.3.2 Continuous Endpoints

Secondary endpoints with continuous values are defined as below:

8.6.3.2.1 Change from baseline in 5D-itch scale and Visual Analog Score (VAS) through week 72

Pruritus will be monitored in this study at day 1, 7, weeks 2, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 56, 64, 72 and EOS.

Observed values, change from baseline, and percentage change from baseline for VAS scores and each dimension of 5D-itch scale and overall score will be summarized using an 8-number summary by visit and treatment group.

Change from baseline for VAS and 5D-itch scale will be analyzed using a restricted maximum likelihood-based mixed model repeated measures (MMRM) technique. The model will include treatment, visit, treatment-by-visit interaction as fixed effects along with baseline NAS score and baseline score for VAS and 5D-itch scale as covariate respectively. Variance-covariance structure to be used to model the within-patient error and results to be presented are described in Section 8.6.1. The details for calculation of score is given in Appendix B.

A by-subject listing will be provided for observed and derived variables for VAS and 5D-itch scale.

8.6.3.2.2 Change from baseline in percentage of fat in the liver as assessed by magnetic resonance imaging proton density fat fraction (MRI-PDFF) at week 12 and week 72

Percentage of fat in the liver assessed by MRI-PDFF at screening, week 12 and week 72, change from baseline and percentage change from baseline will be summarized using an 8-number summary by visit and treatment group.

Change from baseline in percentage of fat in the liver at week 12 and week 72 will be analysed using an analysis of covariance (ANCOVA) model with treatment, baseline diabetes status and use of Vitamin E and/or pioglitazone as fixed effect along with baseline NAS score and baseline MRI-PDFF value as covariate. Least-squares means (using type III sum-of-squares), standard errors, p-value and 2-sided 95% confidence interval for treatment differences will be presented for week 12 and week 72. Significance tests will be based on least-squares means using a two-sided $\alpha = 0.05$.

A by-subject listing will be provided for observed and derived variables.

8.6.3.2.3 Change from baseline in liver stiffness as assessed by magnetic resonance elastograph (MRE) at week 12 and week 72

The liver stiffness assessed by MRE at screening, week 12 and week 72, change from baseline and percentage change from baseline will be summarized using an 8-number summary by visit and treatment group.

Change from baseline in liver stiffness at week 12 and week 72 will be analysed using an analysis of covariance (ANCOVA) model with treatment, baseline diabetes status and use

of Vitamin E and/or pioglitazone as fixed effect along with baseline NAS score and baseline MRE value as covariate. Least-squares means (using type III sum-of-squares), standard errors, p-value and 2-sided 95% confidence interval for treatment differences will be presented for week 12 and week 72. Significance tests will be based on least-squares means using a two-sided $\alpha = 0.05$.

A by-subject listing will be provided for observed and derived variables.

8.6.3.2.4 Change from baseline in TG, total cholesterol (TC), HDL cholesterol, LDL cholesterol and adiponectin through week 72

The parameters, TG, TC, HDL cholesterol, LDL cholesterol and adiponectin are captured in lipid profile panel. All these parameters will be monitored in this study at day 1, weeks 2, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 56, 64, 72 and EOS.

Observed values, change from baseline, and percentage change from baseline for the parameters triglyceride (TG), total cholesterol (TC), HDL cholesterol, LDL cholesterol and adiponectin will be summarized using an 8-number summary by visit and treatment group.

Change from baseline for each of the above parameters will be analyzed separately using a restricted maximum likelihood-based mixed model repeated measures (MMRM) technique. The model will include treatment, visit, treatment-by-visit interaction as fixed effects along with baseline NAS score and baseline value of the respective parameters as covariate. Variance-covariance structure to be used to model the within-patient error and results to be presented are described in Section 8.6.1.

A by-subject listing will be provided for observed and derived variables.

8.6.3.2.5 Pharmacokinetic concentrations of EDP-305 (and metabolites)

PK samples will be collected at day 1, weeks 2, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 56, 64, and 72. On Day 1 and weeks 12, 36, and 72.

The detailed information related to pharmacokinetics analysis is provided in Section 8.10.

[redacted]

[redacted]

A by-subject listing will be provided for Quality of Life Assessment.

[redacted]

[redacted]

8.7 Safety Analyses

All safety analyzes will be based on the safety population and will be performed for all safety variables specified below.

No statistical tests will be performed.

8.7.1 Adverse events

All Adverse Events (AEs) will be classified by Primary System Organ Class (SOC) and Preferred Term (PT) according to the Medical Dictionary for Regulatory Activities (MedDRA) Version 23.0 or higher. All subjects in the safety analysis set will be included in the summaries.

AEs will be classified as pre-treatment AEs and treatment emergent adverse events (TEAEs) and are defined as follows:

Pre-treatment AE: A pre-treatment event is any event that meets the criteria for an AE/SAE and occurs after the subject signs the ICF but before receiving the first administration of study drug.

TEAE: A TEAE is defined as an AE occurring or worsening on or after the first dose of study drug until the last dose of study drug.

Relationship of each AE to study drug:

Related: There is an association between the event and the administration of study drug, a plausible mechanism for the event to be related to the study drug and causes other than the study drug have been ruled out, and/or the event re-appeared on re-exposure to the study drug.

Possibly Related: There is an association between the event and the administration of the study drug and there is a plausible mechanism for the event to be related to study drug, but there may also be alternative etiology, such as characteristics of the subject's clinical status or underlying disease.

Unlikely Related: The event is unlikely to be related to the study drug and likely to be related to factors other than study drug.

Not Related: The event is related to an etiology other than the study drug (the alternative etiology must be documented in the study subject's medical record).

Grade AEs:

Grade AEs (serious and non-serious) in accordance with the NCI/CTCAE scale as presented below. Refer the https://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_8.5x11.pdf provided in reference for the same:

- **Mild** (Grade 1) asymptomatic or mild symptoms; clinical or diagnostic observations only; intervention not indicated
- **Moderate** (Grade 2) minimal, local or noninvasive intervention indicated; limiting age-appropriate instrumental activities of daily living
- **Severe** (Grade 3) Severe or medically significant but not immediately life threatening; hospitalization or prolongation of hospitalization indicated; disabling; limiting self-care activities of daily living
- **Life-threatening** (Grade 4) Life-threatening consequences; urgent intervention indicated
- **Death** (Grade 5) Death related to the AE.

Details for imputing missing or partial start dates of adverse events are described in Section 7.2.2. Imputed Adverse Event dates will be used for determining treatment-emergence.

Summaries of AEs will include the following:

- TEAEs
- Related TEAEs
- Maximum Severity TEAEs
- TEAEs by severity
- TEAEs leading to study drug discontinuation
- TEAEs leading to study discontinuation
- AEs leading to death
- SAEs
- Related Treatment-Emergent SAEs

All TEAEs will be summarized by SOC, PT and treatment group using frequency counts and percentages. In addition, an overall summary for the categories above will be prepared by treatment group and overall.

On-treatment AEs will be summarized separately from those events that occurred during treatment-free follow-up. Those events that occurred during treatment-free follow-up will be summarized by randomized treatment group. The denominator for these events will be only those subjects who entered into the 4 week treatment free follow-up period.

Where a subject has the same adverse event, based on preferred terminology, reported multiple times in the treatment period, the subject will only be counted once at the preferred terminology level in adverse event frequency tables.

Where a subject has multiple adverse events within the same system organ class in the treatment period, the subject will only be counted once at the system organ class level in adverse event frequency tables.

When reporting adverse events by severity, in addition to providing a summary table based on the event selection criteria detailed above, summary table will also be provided based on the most severe event during the treatment period - independent of relationship to study treatment.

8.7.2 Lab Based Stopping Rules

Summaries will be provided using counts and percentages for subjects who meet the stopping rules as described below:

- If ALT or AST increases to >5× Reference baseline (and is at least > 5x ULN)
- If ALT or AST increases to >2× Reference baseline (and is at least > 3x ULN) AND the increase is accompanied by
  - a concomitant total bilirubin increase to >2× Reference baseline (and at least > 2x ULN) OR
  - the INR is concomitantly > 1.5
- If ALT or AST increases to >2× Reference baseline (and is at least > 3x ULN) AND
  - elevations of ALT/AST are accompanied by signs or symptoms of right upper quadrant abdominal pain, anorexia, nausea, vomiting, fever, eosinophilia (> upper limit of normal of absolute eosinophilia count), and/or rash.

8.7.2.1 Close Monitoring for Drug Discontinuation due to Elevated ALT/AST

The following close observation guidelines applies to subjects that repeat assessments show persistent elevations of transaminases, but who do not meet drug discontinuation criteria, and for subjects who discontinue study drug due to ALT/AST elevations.

Close monitoring activities will be initiated:

- If ALT or AST increases > 2x Reference baseline (and is at least > 3x ULN) OR
- If ALT or AST increases > 8x ULN (whichever comes first)

Details related close monitoring is available in protocol section 10.2.

A listing will be used to track and evaluate subjects that are part of the close observation guidelines.

8.7.3 Clinical laboratory evaluations

Summaries of clinical laboratory values and change from baseline will be performed using an 8-number summary by visit and treatment group. All subjects in the safety population

will be included in these summaries. Baseline is defined as the last value collected prior to first dose of study drug.

The number and percentage of subjects with treatment-emergent laboratory abnormalities until the last dose of study drug will be summarized by treatment group. In addition, shift from baseline tables will be generated by visit and treatment group.

For categorical tests: Treatment-emergent abnormal is defined as a change from normal at baseline to abnormal at any post-baseline visit

For continuous tests:

- Treatment-emergent high is defined as a change from a result less than or equal to the high limit at baseline to a value greater than the high limit at any time post-baseline.
  Results will be reported according to any value greater than the high limit, any value greater than 2x ULN and 3x ULN
- Treatment-emergent low is defined as a change from a result greater than or equal to the low limit at baseline to a value less than the low limit at any time post-baseline
  Results will be reported according to any value less than the lower limit, any value less than 2x LLN and 3x LLN.

All summary tables will include the abnormalities that occur during the treatment-free follow-up after the date of last dose. The denominator for these will include any subject who entered into the 4 week treatment-free follow-up phase.

All laboratory data will be included in the data listings and all test values outside the normal range will be flagged as low or high.

Separate listings will be provided for: Pregnancy test, FSH, HIV, HCV, and HBV along with biomarkers for NASH (Cytokeratin (CK)18, GLP-1 (glucagon-like peptide-1)).

8.7.4 Vital signs

Vital signs will include SBP (mmHg), DBP (mmHg), RR (breaths/min), HR (beats/min), and body temperature (°C). Observed values, change from baseline and percent change from baseline for vital signs will be summarized using an 8-number summary by visit and treatment group. In addition, the number and percentage of subjects with significant changes (per the criteria below) in vital signs from baseline will be summarized by treatment.

Criteria for clinically significant changes in vital signs parameters.

Pulse Rate

- < 50 bpm
- >120 bpm
- 30 bpm increase from baseline

- 30 bpm decrease from baseline

Blood Pressure

- SBP > 150 mmHg or DBP > 100 mmHg
- SBP > 200 mmHg or DBP > 110 mmHg

Respiration Rate

- < 8 breaths/min
- 40 breaths/min

Temperature

- 38.3°C
- 1.1°C increase from baseline (Baseline > 36.8°C)

Change in Weight

- 5% increase from baseline
- 5% decrease from baseline

A by-subject listing will be provided.

8.7.5 Physical examinations

A full physical examination will be conducted at screening, week 12 and EOS as indicated in the schedule of assessments and will include a review of the following systems: head/neck/thyroid; eyes/ears/nose/throat (EENT); respiratory; cardiovascular; chest; lymph nodes; abdomen; skin; musculoskeletal; Neurological; Breast; anorectal, and genital examinations will be performed when medically indicated.

Physical examination data will be provided in data listings.

8.7.6 Electrocardiograms (ECG)

12-lead ECG measurements include (Heart Rate (bpm), QRS Duration (msec), PR Interval (msec), QT Interval (msec), RR Interval (msec) and QTcF (msec)) and will be summarized using an 8-number summary by visit and treatment for all safety subjects for observed values and changes from baseline. The overall ECG interpretation (Normal, Abnormal Not Clinically Significant, Abnormal Clinically Significant) will be summarized by presenting the number and percentage of subjects for each treatment group and time-point.

In addition, the number and percentage of subjects meeting or exceeding the following categories will be summarized by treatment group.

- Absolute QT/QTc interval prolongation
    - QTc interval > 450
    - QTc interval > 480
    - QTc interval > 500

- Change from baseline in QTc interval
  - QTc interval increases from baseline >30
  - QTc interval increase from baseline >60

A by-subject listing will be provided.

8.8 Pharmacokinetics (PK) and Pharmacokinetics/Pharmacodynamics (PK/PD) Analysis

During the study, PK samples and PD samples for FGF19, C4 and bile acid will be collected as shown in the schedule of assessments. On day 1 and weeks 12, 36, and 72, a sample will be collected predose and two samples postdose; the first postdose sample will be collected 1 to 3 hours postdose and the second postdose sample at least 1 hour later. At all other visits, PK samples will be collected predose.

8.8.1 Pharmacokinetic Analyses

The PK analysis will use the pharmacokinetic population and will include only subjects in active treatment arms.

The concentration data for EDP-305 and each metabolite collected according to schedule of events and will be summarized by active treatment arm, visit, and nominal time point. Concentrations that are below the limit of quantitation (BLQ) will be treated as zero for the computation of descriptive statistics. If more than 50% of subjects have postdose concentration values below BLQ, descriptive statistics will not be presented except for maximum and BLQ will be displayed for mean and minimum. The number of observations, arithmetic mean, standard deviation (SD), % coefficient of variation (%CV), median, minimum, maximum, geometric mean, and %CV of the geometric mean (%GCV) will be displayed.

Individual data will be presented in listing.

8.8.2 Pharmacodynamic Analyses

The PD analysis will be based on subjects' randomized treatment assignment using all subjects belonging to the ITT population with available dosing data. This will include all treatment arms (active and placebo).

A descriptive summary will be provided for the FGF19, C4, and bile acid endpoints including measured concentrations, change from baseline, and percentage change from baseline using the 8-number summary by visit, treatment group, and bin time midpoint.

A by-subject listing will be provided for observed and derived variables.

8.8.3 Pharmacokinetics/Pharmacodynamics (PK/PD) Analysis

Scatter plots will be created using all subjects in the PK population. Plots for percentage change from baseline in each PD activity with EDP-305 concentration value will be generated for week 12 pre-dose.

Additionally the following PK/PD plots will be created using all subjects in the PK population. Other PD plots will be created using all subjects in the efficacy population,

including placebo, as applicable. The following scatterplots will be generated:

- Percentage change from Baseline in ALT with the percentage change from Baseline in FGF19 (Predose Week 12)
- Percentage change from Baseline in ALT with the percentage change from Baseline in C4 (Predose Week 12)
- Percentage change from Baseline in ALT with the percentage change from Baseline in Bile Acid (Predose Week 12)
- Percentage change from Baseline in ALT with EDP-305 concentration value (predose at Week 12)
- Percentage change from Baseline in GGT with the percentage change from Baseline in FGF19 (Predose Week 12)
- Percentage change from Baseline in GGT with the percentage change from Baseline in C4 (Predose Week 12)
- Percentage change from Baseline in GGT with the percentage change from Baseline in Bile Acid (Predose Week 12)
- Percentage change from Baseline in GGT with EDP-305 concentration value (Predose Week 12)
- Percentage change from Baseline in Alkaline Phosphatase with the percentage change from Baseline in FGF19 (Predose Week 12)
- Percentage change from Baseline in Alkaline Phosphatase with the percentage change from Baseline in C4 (Predose Week 12)
- Percentage change from Baseline in Alkaline Phosphatase with the percentage change from Baseline in Bile Acid (Predose Week 12)
- Percentage change from Baseline in Alkaline Phosphatase with EDP-305 concentration value (Predose Week 12)
- Percentage change from Baseline in MRI-PDFF with the percentage change from Baseline in FGF19 (Predose Week 12)
- Percentage change from Baseline in MRI-PDFF with the percentage change from Baseline in C4 (Predose Week 12)
- Percentage change from Baseline in MRI-PDFF with the percentage change from Baseline in Bile Acid (Predose Week 12)
- Percentage change from Baseline in MRI-PDFF with EDP-305 concentration value (Predose Week 12)

- Percentage change from Baseline in MRE with the percentage change from Baseline in FGF19 (Predose Week 12)
- Percentage change from Baseline in MRE with the percentage change from Baseline in C4 (Predose Week 12)
- Percentage change from Baseline in MRE with the percentage change from Baseline in Bile Acid (Predose Week 12)
- Percentage change from Baseline in MRE with EDP-305 concentration value (Predose Week 12)
- Percentage change from Baseline in ALT with the percentage change from Baseline in FGF19 (Predose Week 72)
- Percentage change from Baseline in ALT with the percentage change from Baseline in C4 (Predose Week 72)
- Percentage change from Baseline in ALT with the percentage change from Baseline in Bile Acid (Predose Week 72)
- Percentage change from Baseline in ALT with EDP-305 concentration value (predose at Week 72)
- Percentage change from Baseline in GGT with the percentage change from Baseline in FGF19 (Predose Week 72)
- Percentage change from Baseline in GGT with the percentage change from Baseline in C4 (Predose Week 72)
- Percentage change from Baseline in GGT with the percentage change from Baseline in Bile Acid (Predose Week 72)
- Percentage change from Baseline in GGT with EDP-305 concentration value (Predose Week 72)
- Percentage change from Baseline in Alkaline Phosphatase with the percentage change from Baseline in FGF19 (Predose Week 72)
- Percentage change from Baseline in Alkaline Phosphatase with the percentage change from Baseline in C4 (Predose Week 72)
- Percentage change from Baseline in Alkaline Phosphatase with the percentage change from Baseline in Bile Acid (Predose Week 72)
- Percentage change from Baseline in Alkaline Phosphatase with EDP-305 concentration value (Predose Week 72)

- Percentage change from Baseline in MRI-PDFF with the percentage change from Baseline in FGF19 (Predose Week 72)
- Percentage change from Baseline in MRI-PDFF with the percentage change from Baseline in C4 (Predose Week 72)
- Percentage change from Baseline in MRI-PDFF with the percentage change from Baseline in Bile Acid (Predose Week 72)
- Percentage change from Baseline in MRI-PDFF with EDP-305 concentration value (Predose Week 72)
- Percentage change from Baseline in MRE with the percentage change from Baseline in FGF19 (Predose Week 72)
- Percentage change from Baseline in MRE with the percentage change from Baseline in C4 (Predose Week 72)
- Percentage change from Baseline in MRE with the percentage change from Baseline in Bile Acid (Predose Week 72)
- Percentage change from Baseline in MRE with EDP-305 concentration value (Predose Week 72)

Box plot of EDP-305 and EP-022679 concentration (last predose) for subjects with adverse events occurring in at least 5% of the combined active subjects.

Lastly, the analyses below will be carried out to provide additional data for better understanding of specific clinical events and manifestations during the course of the study. These analyses will be conducted for subjects with or without pruritus:

PK and pruritus:

- Summary of plasma concentration of EDP-305 and EP-022679 per time point and treatment group by visit for subjects with and without pruritus
- Summary of PK parameters by treatment group by visit for subjects with and without pruritus
- Boxplot of predose concentration at Week 2,4,8 and 12 for EDP-305 and EP-022679 by treatment group for subjects with and without pruritus
- Listing of EDP-305 and EP-022679 concentration value for subjects with pruritus

C4 and pruritus:

- Summary of C4, change from baseline and percentage change from baseline by hour by visit for subjects with and without pruritus
- Summary of AUC_{0-8} and AUC_{2-8} for C4, change from baseline and percentage change from baseline at Week 12 for subjects with and without pruritus

8.9 Subgroup Analysis

Subgroup analyses will be performed for the primary endpoint and the secondary endpoints to determine whether significant differences exist in primary and secondary endpoint results between subgroups.

These subgroup analyses will be carried out using the subjects from the ITT and PP population.

The list of potential subgroups (with applicable definitions in parentheses) includes the following:

- Baseline Fibrosis score (2, 3),
- MRE week 72 response: (<15% reduction vs ≥15% reduction in percentage of change from baseline),
- MRI-PDFF week 72 response: (<30% reduction vs ≥30% reduction in percentage of change from baseline),
- Both MRE week 72 response (≥15% reduction in percentage of change from baseline) and MRI-PDFF week 72 response (≥30% reduction in percentage of change from baseline) vs non-response.

A logistic regression model will be used for categorical endpoints. The model will include the treatment group, diabetes status, use of Vitamin E and/or pioglitazone, subgroup and subgroup-by-treatment interaction as fixed effects and baseline NAS score as covariate. Odds ratios and 95% confidence intervals will be reported.

For continuous endpoints an ANCOVA model with treatment group, diabetes status, use of Vitamin E and/or pioglitazone, subgroup and subgroup-by-treatment interaction as fixed effects and baseline NAS along with baseline parameter of interest as covariates. Least-squares means (using type III sum-of-squares), standard errors, p-value and 2-sided 95% confidence interval for treatment differences will be reported.

For continuous endpoints that are repeated over visits a MMRM technique will be used. The model will include treatment, visit, subgroup and interaction effect of treatment, visit and subgroup as fixed effects and baseline NAS along with baseline parameter of interest as covariates. Least-squares means (using type III sum-of-squares), standard errors, p-value and 2-sided 95% confidence interval for treatment differences will be presented for each visit

Forest plots will be presented for the sub group analysis mentioned above to describe the association using ITT and PP population.

8.10 Interim Analysis

Two partially unblinded interim analyses (IA) are planned for this study. To protect study integrity, the IA will be performed by an independent statistician and statistical reporting group who are not involved in the clinical trial conduct and are not responsible for the final analysis of clinical trial data. Each interim analysis will be governed by a separate interim analysis charter.

8.10.1 Interim Analysis 1 (IA1)

The first interim analysis will be conducted when approximately 75 subjects have completed the week 12 visit to review safety, tolerability. The key secondary endpoints in IA1 will be noninvasive tests (NITs): MRI-PDFF, MRE and non-invasive fibrosis markers. No analysis of the primary endpoint will be performed in IA1. PK, PD and PK/PD endpoints will be summarized as described in Section 8.8.

The results to be submitted in IA1 will include all the displays for safety endpoints: treatment compliance, AEs, lab, ECG, vital signs, concomitant medications and physical examination; disposition and demographics results; and results for key secondary endpoints: MRI-PDFF, MRE and non-invasive fibrosis markers.

Analysis for the liver MRI-PDFF, liver MRE and non-invasive markers is provided in Section 8.6.3.2.2, Section 8.6.3.2.3 and Section 8.6.4.1 respectively.

8.10.2 Interim Analysis 2 (IA2)

The second interim analysis when approximately 35-40% of subjects have completed the week 72 biopsy. The analysis for the primary endpoint will be performed in IA2. A Lan-Demets alpha spending function with O'Brien-Fleming boundaries will be applied to control type I error at 0.025 significance level (two-sided) for each primary endpoint. Adjusted alpha obtained with this method will be used in second interim analysis.

Each primary endpoint will be analysis using the logistic regression model given in Section 8.6.2. A step-down sequential testing procedure will be used to control multiplicity in testing each dose vs placebo.

PK, PD and PKPD endpoints will be summarized as described in Section 8.8.

8.11 COVID-19 Impact on Study

Additional analysis will be performed to recognize and indicate the impact of participation of subjects by COVID-19. The impact of COVID-19 will be summarized by treatment group and overall by visit for all screened subjects. This table will present the subjects affected by COVID-19 in all aspects (e.g., missed visit and/or assessment due to COVID-19, remote visit due to COVID-19, modified assessment due COVID-19). In addition important protocol deviations related to COVID-19 using randomized subjects will be summarized.

A by-subject listing will also be provided for all subjects that have their participation impacted by COVID-19.

9 CHANGES TO PLANNED ANALYSIS FROM STUDY PROTOCOL

This study was terminated by Enanta on October 4th, 2021. At the time the study was terminated no subject had completed Week 72 study assessments. Any analyses planned through week 72 will present all data collected through study termination, and only summary statistics will be provided post Week 12 visit. Any analyses at week 72 will not be performed.

Consequently, the following analyses will not be performed,

SAP Section	Analysis: Endpoint(s)
8.6.2	Primary Efficacy: All
8.2.6.1	Sensitivity Analyses for Primary Endpoints: All
8.6.3.1	Secondary Efficacy Analysis: Categorical Endpoints (Responder at Week 72) – All
8.6.2.3.4	Secondary Efficacy Analysis: Continuous Endpoint Adiponectin through week 72
8.6.4.2	Exploratory Analysis at Week 72: NAFLD Fibrosis Score, AST to Platelet Ratio Index
8.6.4.3	Exploratory Analysis at Week 72: CRP, IL-6, IL-1β, TNF-α, TNF-β, alpha2 macroglobulin and haptoglobin levels
8.6.4.5	Exploratory Analysis: ASCVD Score
8.8	Pharmacokinetics (PK) and Pharmacokinetics/Pharmacodynamics (PK/PD) Analysis: All Analyses with the following exceptions: • Summary of AUC_{0-8} and AUC_{2-8} for C4, change from baseline and percentage change from baseline at Week 12 for subjects with and without pruritus • Scatterplot of FGF19 concentrations at Week 12 vs FGF19 Week 12 Change from Baseline value • Scatterplot of C4 concentrations at Week 12 vs C4 Week 12 Change from Baseline value • Scatterplot of Bile Acid concentrations at Week 12 vs Bile Acid Week 12 Change from Baseline value • Boxplot of predose concentration at Weeks 2, 4, 8, and 12 for EDP-305 and EP-022679 by treatment group for subjects with and without pruritus • Listing of EDP-305 & EP-022679 concentrations for subjects with pruritus
8.9	Subgroup Analyses: All
8.10.2	Interim Analysis #2: All
8.11	Analysis of COVID-19 Impact on Study: All

- The following listings have been added: Informed Consent and Rescreening data will be listed for all enrolled subjects.
- Inclusion and exclusion criteria will be listed for all enrolled subjects.

10 REFERENCES

1. Clinical Study Protocol (Amendment 1.0) Phase 2b Randomized, Double Blind, Placebo-Controlled, Multicenter Study Evaluating Safety And Efficacy Of Edp-305 In Subjects With Liver-Biopsy Proven Non-Alcoholic Steatohepatitis (Nash) (Argon-2)
2. Reboussin, D. M., DeMets, D. L., Kim, K. M., & Lan, K. K. (2000). Computations for group sequential boundaries using the Lan-DeMets spending function method. Control Clin Trials, 21(3), 190-207.
3. Elman, S., Hynan, L. S., Gabriel, V., & Mayo, M. J. (2010). The 5-D itch scale: a new measure of pruritus. Br J Dermatol, 162(3), 587-593. doi:10.1111/j.1365-2133.2009.09586.x
4. ICH Topic E9: Statistical Principles for Clinical Trials (CPMP/ICH/363/96 – adopted March 1998).
5. https://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_8.5x11.pdf
6. Agopian, V. G., Kaldas, F. M., Hong, J. C., Whittaker, M., Holt, C., Rana, A., . . . Busuttil, R. W. (2012). Liver transplantation for nonalcoholic steatohepatitis: the new epidemic. Ann Surg, 256(4), 624-633. doi:10.1097/SLA.0b013e31826b4b7e.

11 APPENDICES

Appendix A – Statistical Model

SAS code for Logistic regression analysis

```
PROC GENMOD data=dataset ;
    class trt01pn(ref=0) diab_st vitE_pio;
    freq count;
    model Resp = trt01pn diab_st vitE_pio baseline_NAS / alpha=0.xx dist=binomial
    link=logit ;
    lsmeans trt01pn/ CL pdiff exp;
    ods output      lsmeans = lsmeans1
                    diffs=diffs1;
RUN;
```

Note:

1. Use data without any imputation.
2. Resp: Responder (Yes/No)
3. diab_st: Baseline diabetes status
4. vitE_pio: Use of Vitamin E and/or pioglitazone
5. Baseline_NAS: Baseline value of NAS
6. 0.xx=level of significance to be used
7. Use numeric variables for treatment group
8. Code **trt01pn** as Placebo=0, EDP 305 2 mg=1, EDP 305 1.5 mg=2

For Primary analysis: The EDP-305 2 mg will be compared to placebo using a two-sided alpha of 0.0xx (adjusted alpha/2) first then if the difference is statistically significant, EDP-305 1.5 mg will be compared to placebo using a two-sided alpha of 0.0xx (adjusted alpha/2).

SAS code for Fisher Exact Test

```
ods output fishersexact=fishers;

PROC FREQ data=dataset;

tables trt01pn*responder /exact out=count;

RUN;
```

Note:

1. Use data without any imputation
2. Fishers: Output dataset for Fisher's Exact test (for p-value)

SAS code for ANCOVA analysis

```
PROC MIXED data = dataset method=type3;
    class trt01pn(ref=0)  diab_st vitE_pio;
    model CHG = trt01pn diab_st vitE_pio baseline_NAS baseline_parameter /ddfm=KR;
    lsmeans trt01pn /pdiff ;
    ods output      LSMEANS=lsm
                    DIFFS=diff ;
RUN;
```

Note:

1. Use data without any imputation.
2. CHG: Change from baseline.
3. diab_st: Baseline diabetes status
4. vitE_pio: Use of Vitamin E and/or pioglitazone
5. Baseline_NAS: Baseline value of NAS
6. Baseline_parameter: Baseline value of the parameter to be analysed.
7. Use numeric variables for treatment group
8. Code **trt01pn** as Placebo=0, EDP 305 2 mg=1, EDP 305 1.5 mg=2

SAS code for MMRM analysis

```
PROC MIXED data = dataset method=type3;
    class trt01pn(ref=0)  visitn usubjid ;
    model CHG = trt01pn visitn trt01pn*visitn baseline_NAS baseline_parameter
    /ddfm=KR;
    repeated trt01pn / type = UN subject = usubjid;
    lsmeans trt01pn*avisitn /pdiff ;
    ods output      LSMEANS=lsm
                    DIFFS=diff
RUN;
```

Note:

1. Use data without any imputation.
2. CHG: Change from baseline.
3. Baseline_NAS: Baseline value of NAS
4. Baseline_parameter: Baseline value of the parameter to be analysed.
5. Use numeric variables for visits and treatment group
6. Code **trt01pn** as Placebo=0, EDP 305 2 mg=1, EDP 305 1.5 mg=2
7. If model does not converge then use type=TOEPH

Appendix B – Scoring for Scales

5D-itch scale

5D-itch scale is developed by Elman (Elman, Hynan, Gabriel, & Mayo, 2010) and will be used to assess pruritus in subjects. The five dimensions (aka domains) assessed in 5D-itch scale are duration, degree, direction, disability, and distribution.

The disability domain will be the highest score obtained from the daily activities of sleep, leisure/social activities, housework/errands, and work/school. The disability domain will only be calculated if at least 3 daily activities are documented.

The distribution domain includes 16 locations of itch and is the categorized to a scale of 0 to 5, based on the sum of the number of affected locations: sum of 0 to 2 = score of 1, sum of 3 to 5 = score of 2, sum of 6 to 10 = score of 3, sum of 11 to 13 = score of 4, sum of 14 to 16 = score of 5.

The total 5D score is obtained by summing up the five domain scores and ranges between 5 (no pruritus) and 25 (most severe pruritus). The total score will not be calculated if any of the domain scores is missing.

Quality of Life scales - CLDQ-NASH

CLDQ-NASH scale includes 36 items grouped into 6 domains: Abdominal Symptoms, Activity/Energy, Emotional Health, Fatigue, Systemic Symptoms, and Worry. This questionnaire is designed to find out how a subject had been feeling during the last two weeks Subject will be asked about their symptoms related to their fatty liver disease, how they have been affected in doing activities, and how their mood have been.

In CLDQ-NASH, patients are asked about how frequently they experience certain problems or how they have been affected in doing activities impairing various aspects of their well-being; a 1-7 Likert scale is used for the responses such that the score of 1 corresponds to a problem is experienced “All of the time”, and the score of 7 to “None of the time”.

The scores will be calculated separately for each domain as an average of the domain’s items. In all domains, greater scores reflect better health. The total CLDQ-NASH score will be obtained as the average of the domain scores.

Cardiovascular risk score (ASCVD)

Data Derivation Details to Obtain Scale Scores for SF-36

SF-36 questionnaire includes 8 domains with 36 items. This questionnaire designed to get idea about the subjects views about his/her health and to track of how the subject is feeling and if subject is able to do your usual activities.

	DERIVATION
SF-36 PF scale score	raw score = sum (items 3A, 3B, 3C, 3D, 3E, 3F, 3G, 3H, 3I, 3J) PF = [(raw score -10)/20] * 100 PF_Z = (PF – 83.29094) / 23.75883 PF scale score = (PF_Z*10) + 50 When calculating the raw score, if 5 or more of the items are non-missing then replace any missing values as follows: Calculate the mean of the answered questions, using the number of answered questions as the denominator. Substitute the mean for any missing scores. Otherwise, if less than 5 of the items are non-missing then PF scale score is missing. The response scale for each activity ranges from 1 to 3 where 1=limited a lot, 2=limited a little, and 3=not limited at all. A higher PF scale score indicates better physical functioning.

	DERIVATION
SF-36 RP scale score	raw score = sum (items 4A, 4B, 4C, and 4D) RP = [(raw score -4)/16] * 100 RP_Z = (RP – 82.50964) / 25.52028 RP scale score = (RP_Z * 10) + 50 When calculating the raw score, if 2 or more of the items are non-missing then replace any missing values as follows: Calculate the mean of the answered questions, using the number of answered questions as the denominator. Substitute the mean for any missing scores. Otherwise, if less than 2 of the items are non-missing then RP scale score is missing. The response scale for each item ranges from 1 to 5 where 1=All of the time, 2=Most of the time, 3=Some of the time, 4=A little of the time, and 5=None of the time. A higher RP scale score indicates better role-physical functioning.

SF-36 BP scale score	raw score = sum (reversed item 7 and reversed item 8) BP = [(raw score -2) / 10]*100 BP_Z = (BP – 71.32527) / 23.66224 BP scale score = (BP_Z * 10) + 50 Reverse direction of Item 7 as follows if =1, set to 6 if =2, set to 5.4 if =3, set to 4.2 if=4, set to 3.1 if=5, set to 2.2 if=6, set to 1 Reverse direction of item 8 as follows: if =1 and original value of item 7=1, set to 6 if =1 and original value of item 7>=2, set to 5 if=2, set to 4 if=3, set to 3 if=4, set to 2 if=5, set to 1 If item 7 is answered and item 8 is missing, set 8 = reversed 7 as defined above. If 8 is answered and 7 is missing, set 7 as reverse item 8 as follows if =1, set to 6 if =2, set to 4.75 if=3, set to 3.5 if=4, set to 2.25 if=5, set to 1 If 1 or more questions were answered, calculate BP scale score as defined above. If neither question was answered then BP scale score is missing. The scale for Question 7, amount of bodily pain, ranges from 1 to 6 where 1=None, 2=Very mild, 3=mild, 4=Moderate, 5=Severe, and 6=Very severe. The scale for Question 8, the degree to which pain interfered with normal work, ranges from 1 to 5 where 1=Not at all, 2=A little bit, 3=Moderately, 4=Quite a bit, and 5=Extremely. A higher BP scale score indicates lack of bodily pain.

	DERIVATION
SF-36 GH scale score	raw score = sum (reversed item 1, item 11A, reversed 11B, 11C and reversed 11D) GH = [(raw score -5) / 20]*100 GH_Z = (GH – 70.84570) / 20.97821 GH scale score = (GH_Z * 10) + 50 Reverse direction of Item 1 as follows: if =1, set to 5 if =2, set to 4.4 if =3, set to 3.4 if =4, set to 2 if =5, set to 1 Reverse direction of item 11B and 11D by subtracting score from 6. When calculating the raw score, if 3 or more of the items are non-missing then replace any missing values as follows: Calculate the mean of the answered questions, using the number of answered questions as the denominator. Substitute the mean for any missing scores. Otherwise, if less than 3 of the items are non-missing then GH scale score is missing. Responses for Question 1, an assessment of self-perceived health status, range from 1 to 5 where 1=Excellent, 2=Very good, 3=Good, 4=Fair, and 5=Poor. Responses for the items in Question 11 range from 1 to 5 where 1=Definitely true, 2=Mostly true, 3=Don't know, 4=Mostly false, and 5=Definitely false and reflect the subject's perception of their relative health and expectations of their future health status. A higher GH scale score indicates better general health perceptions.

	DERIVATION
SF-36 VT scale score	raw score = sum (reversed item 9a, reversed 9e, 9g and 9i) VT = [(raw score -4)/16] * 100 VT_Z = (VT – 58.31411) / 20.01923 VT scale score = (VT_Z * 10) + 50 Reverse direction of Items 9a and 9e by subtracting score from 6. When calculating the raw score, if 2 or more of the items are non-missing then replace any missing values as follows: Calculate the mean of the answered questions, using the number of answered questions as the denominator. Substitute the mean for any missing scores. Otherwise, if less than 2 of the items are non-missing then VT scale score is missing.
SF-36 SF scale score	raw score = sum (reversed 6 and 10) SF = [(raw score -2) / 8] * 100 SF_Z = (SF – 84.30250) / 22.91921 SF scale score = (SF_Z * 10) + 50 Reverse direction of score for item 6 by subtracting score from 6. When calculating the raw score, if 1 of the items is missing then substitute the missing score with the score on the non-missing item. If both items are missing then SF scale score is missing. Responses to Question 6, an assessment of the extent to which health/emotional problems interfered with social activities, range from 1 to 5 where 1=Not at all, 2=Slightly, 3=Moderately, 4=Quite a bit, and 5=Extremely. Responses to Question 10 reflect the amount of time that health/emotional problems interfered with social activities and range from 1 to 5 where 1=All of the time, 2=Most of the time, 3=Some of the time, 4=A little of the time, and 5=None of the time. A higher SF scale score indicates better social functioning.

	DERIVATION
SF-36 RE scale score	raw score = sum (items 5A, 5B, and 5C) RE = [(raw score -3) / 12] * 100 RE_Z = (RE – 87.39733) / 21.43778 RE scale score = (RE_Z * 10) + 50 When calculating the raw score, if 2 or more of the items are non-missing then replace any missing values as follows: Calculate the mean of the answered questions, using the number of answered questions as the denominator. Substitute the mean for any missing scores. Otherwise, if less than 2 of the items are non-missing then RE scale score is missing. Responses to the items in Question 5 range from 1 to 5 where 1=All of the time, 2=Most of the time, 3=Some of the time, 4=A little of the time, and 5=None of the time. A higher RE scale score indicates better role-emotional functioning.

	DERIVATION
SF-36 MH scale score	raw score = sum (items 9B, 9C, reversed 9D, 9F and reversed 9H) MH = [(raw score - 5)/20]*100 MH_Z = (MH – 74.98685) / 17.75604 MH scale score = (MH_Z * 10) + 50 Reverse direction of scores for 9D and 9H, by subtracting score from 6. If 3 or more of the items are non-missing then replace any missing values as follows: Calculate the mean of the answered questions, using the number of answered questions as the denominator. Substitute the mean for any missing scores. Otherwise, if less than 3 of the items are non-missing then MH scale score is missing. The scale for these items ranges from 1 to 5 where 1=All of the time, 2=Most of the time, 3=Some of the time, 4=A little of the time, and 5=None of the time. A higher MH scale score indicates better mental health.
SF-36 TR scale score	raw score = item 2 TR scale score = raw score The scale for this item ranges from 1 to 5 where 1=Much better now than one week ago, 2=Somewhat better now than one week ago, 3=About the same as one week ago, 4=Somewhat worse now than one week ago, and 5=Much worse now than one week ago. A higher TR scale score indicates worse general health currently relative to one week previous.

	DERIVATION
SF-36 PCS score	PCS score includes the 8 scales for GH, PF, RP, RE, SF, MH, BP, and VT. Raw Score = PF_Z *.42402 + RP_Z*.35119 + BP_Z*.31754 + GH_Z*.24954 + VT_Z*.02877 + SF_Z*-.00753 + RE_Z*-.19206 + MH_Z*-.22069 PCS Summary Scale Score = (raw score *10) + 50 Raw Score is missing if one of the component scale scores is missing.
SF-36 MCS score	MCS score includes the 8 scales for GH, PF, RP, RE, SF, MH, BP, and VT. Raw Score = PF_Z *-.22999 + RP_Z*-.12329 + BP_Z*-.09731 + GH_Z*-.01571 + VT_Z*.23534 + SF_Z*.26876 + RE_Z*.43407 + MH_Z*.48581 MCS Summary Concept Score = (raw score *10) + 50 Raw Score is missing if one of the component scale scores is missing.

Appendix C – Analysis Visit Windows

For the purpose of conducting analyses, the visit windows noted below will be utilized for binning study visits. In the event that multiple visits fall within the same window, the scheduled visit will take precedence. If the scheduled visit was not attended, the closest visit to the target within the window will be utilized. If multiple visits are equidistant from the target date, the later of the two visits will be used.

Nominal Visit	Visit Window Lower Bound	Visit Target (Days)	Visit Window Upper Bound
Screening	N/A	N/A	-1
Day 1	1	1	1
Day 7	4	7	10
Week 2	11	14	20
Week 4	21	28	41
Week 8	42	56	69
Week 12	70	84	97
Week 16	98	112	125
Week 20	126	140	153
Week 24	154	168	181
Week 28	182	196	209
Week 32	210	224	237
Week 36	238	252	265
Week 40	266	280	293
Week 44	294	308	321
Week 48	322	336	363
Week 56	364	392	419
Week 64	420	448	475
Week 72	476	504	517
Week 76	518	532	N/A