MEDPACE

STATISTICAL ANALYSIS PLAN

Protocol Title:	A Phase 2, Randomized Double-Blind, Placebo-Controlled, Multi-Center Study to Assess the Efficacy, Safety and Tolerability of Oral LPCN 1144 in Subjects with Nonalcoholic Steatohepatitis (NASH)
Protocol Number:	LPCN 1144-18-002
Protocol Version/Date:	Version 1.0 / 6 June 2019 Version 2.0 / 3 September 2019 Version 3.0 / 7 January 2020
Investigational Product:	LPCN 1144 (Testosterone Undecanoate for NASH)
Sponsor:	Lipocine Inc. 675 Arapeen Drive, Suite 202 Salt Lake City, Utah 84108
SAP Version/Date:	Version 2.0 / 16 March 2021

SIGNATURE PAGE

Protocol Title: A Phase 2, Randomized Double-Blind, Placebo-Controlled, Multi-Center Study to Assess the Efficacy, Safety and Tolerability of Oral LPCN 1144 in Subjects with Nonalcoholic Steatohepatitis (NASH)

Protocol Number: LPCN 1144-18-002

SAP Version/Date: Version 2.0 / 16 March 2021

We, the undersigned, have reviewed and approved this Statistical Analysis Plan:

Signature **Date**

Project Statistician
Medpace, Inc.

Director, Biostatistics
Medpace, Inc.

Medical Director
Medpace, Inc.

Vice President, Product Development
Lipocine Inc.

VERSION HISTORY

Version	Version Date	Description
1.0	30 October 2020	Original signed version
2.0	16 March 2021	Added DXA analysis for Week 20 database lock

TABLE OF CONTENTS

1 Introduction 8
2 Study Overview 8
2.1 Study Objectives 8
2.1.1 Primary Objective 8
2.1.2 Secondary Objectives 8
2.1.3 Safety Objectives 8
2.2 Study Design 8
2.2.1 Overview 8
2.2.2 Randomization and Blinding 9
2.2.3 Study Drug 9
2.2.4 Sample Size Determination 10
2.3 Study Outcome Measures 10
2.3.1 Primary Outcome Measures 10
2.3.2 Key Secondary Outcome Measures 10
2.3.3 Other Secondary Outcome Measures 10
3 Statistical Methodology 11
3.1 General Considerations 11
3.1.1 Analysis Day 11
3.1.2 Analysis Visits 11
3.1.3 Definition of Baseline 12
3.1.4 Summary Statistics 12
3.1.5 Multiple Comparison/Multiplicity 12
3.1.6 Handling of Dropouts and Missing Data 13
3.1.7 The Impact of Coronavirus Disease 2019 (COVID-19) 13
3.2 Analysis Populations 13
3.2.1 Intent-to-Treat (ITT) Analysis Dataset 13
3.2.2 Modified Intent-to-Treat (mITT) Analysis Dataset 13
3.2.3 Per-Protocol (PP) Analysis Datasets 13
3.2.4 Safety Analysis Dataset 14
3.3 Subject Data and Study Conduct 14
3.3.1 Subject Disposition 14
3.3.2 Protocol Deviations 14
3.3.3 Analysis Populations 15
3.3.4 Demographic and Baseline Characteristics 15
3.3.5 Medical History 15
3.3.6 Concomitant Medications 16
3.3.7 Study Drug Exposure and Compliance 16
3.4 Efficacy Assessment 17
3.4.1 Primary Efficacy Outcome Measures 17

3.4.2 Key Secondary Efficacy Outcome Measures 19
3.4.3 Other Secondary Efficacy Outcome Measures....21
3.4.4 Subgroups 26
3.5 Quality of Life Assessment....27
3.5.1 Functional Activity Questionnaire (NHANES III)....28
3.5.2 Health-Related Quality of Life (HR-QOL) measured by Short Form-36 28
3.5.3 Chronic Liver Disease Questionnaire (CLDQ)....28
3.5.4 Patient Global Impression of Severity (PGI-S) 28
3.5.5 Patient Global Impression of Change (PGI-C)....28
3.5.6 Sexual Desire and Distress Questionnaire....29
3.5.7 International Prostate Symptom Score (I-PSS) 29
3.6 Safety Assessment 29
3.6.1 Adverse Events (AEs)....29
3.6.2 Clinical Laboratory Tests 30
3.6.3 Vital Signs....30
3.6.4 Fibroscan....31
3.6.5 Gait Speed....31
3.6.6 Hand Grip Strength Test 31
3.6.7 Cardiovascular Risk Evaluation 31
3.6.8 Other Safety Assessments....31
4 Data and Safety Monitoring Board (DSMB)....31
5 Analysis Timing....32
5.1 Primary Analysis Reviews....32
5.2 Week 20 Analysis Reviews 32
5.3 Interim Analysis 32
6 Changes from Protocol-Specified Statistical Analyses 33
7 Programming Specifications 34

LIST OF ABBREVIATIONS

Abbreviation	Definition
AUDIT	Alcohol Use Disorders Identification Test
AE	Adverse Event
AESI	Adverse Event of Special Interest
ALP	Alkaline Phosphatase
Alpha SMA	Alpha-Smooth Muscle Actin
ALT	Alanine Aminotransferase
ANCOVA	Analysis of Covariance
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Chemical
BID	Twice Daily
BLQ	Below Limit of Quantification
BMI	Body Mass Index
BSAP	Bone Specific Alkaline Phosphatase
BUN	Blood Urea Nitrogen
CBG	Cortisol Binding Globulin
CK	Creatine Kinase
CLDQ	Chronic Liver Disease Questionnaire
COVID-19	Coronavirus Disease 2019
CP	Child Pugh
CRN	Clinical Research Network
CRF	Case Report Form
CRO	Contract Research Organization
CSR	Clinical Study Report
cT1	Corrected T1
CTCAE	Common Terminology Criteria for Adverse Events
DHT	Dihydrotestosterone
DILI	Drug-Induced Liver Injury
DRE	Digital Rectal Examination
DSMB	Data and Safety Monitoring Board
DXA	Dual Energy X-ray Absorptiometry
eDISH	Evaluation of Drug-Induced Serious Hepatotoxicity
ECG	Electrocardiogram
ELF™	Enhanced Liver Fibrosis Panel
EOT	End of Treatment
FAS	Full Analysis Set
FSH	Follicle Stimulating Hormone
GFR	Estimated Glomerular Filtration Rate
GGT	Gamma Glutamyl Transferase
HA	Hyaluronic Acid
HDL-C	High-Density Lipoprotein Cholesterol
HOMA	Homeostasis Model Assessment
HSP47	Heat-Shock Protein 47
HR-QOL	Health-Related Quality of Life
INR	International Normalized Ratio
I-PSS	International-Prostate Symptom Score
ITT	Intent-to-Treat

Abbreviation	Definition
LDL-C	Low-Density Lipoprotein Cholesterol
LH	Luteinizing Hormone
LLOQ	Low Limit of Quantification
LOCF	Last Observation Carried Forward
MedDRA	Medical Dictionary for Regulatory Activities
MAR	Missing at Random
MCP-1	Monocyte Chemotactic Protein-1
MELD	Model for End-Stage Liver Disease
mITT	Modified Intent-to-Treat
MMRM	Mixed-Model Repeated-Measures
MNAR	Missing not at Random
MRI-PDFF	Magnetic Resonance Imaging - Proton Density Fat Fraction
NAB	Neutralizing Antibody
NAFLD	Nonalcoholic Fatty Liver Disease
NAS	NAFLD Activity Score
NASH	Nonalcoholic Steatohepatitis
PIIINP	N-terminal propeptide of type III collagen
PGI-C	Patient Global Impression of Change
PGI-S	Patient Global Impression of Severity
PP	Per Protocol
PT	Prothrombin Time
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SHBG	Sex Hormone Binding Globulin
SF-36	Short Form 36 Health Survey
SOA	Schedule of Activities
SOC	System Organ Class
T	Testosterone
TB	Total Bilirubin
TEAE	Treatment-Emergent Adverse Event
TESAE	Treatment-Emergent Serious Adverse Event
TFLs	Tables, Figures, and Listings
TGF-Beta	Transforming growth Factor-Beta
TIMP-1	Tissue inhibitor of metalloproteinase 1
TNF-alpha	Tissue Necrosis Factor-alpha
TSH	Thyroid Stimulating Hormone
TU	Testosterone Undecanoate
WHO	World Health Organization
VLDL-C	Very Low-Density Lipoprotein Cholesterol

1 INTRODUCTION

The purpose of this Statistical Analysis Plan (SAP) is to provide a description of the statistical methods to be implemented for the analysis of data from the study with protocol number LPCN 1144-18-002. The SAP will be finalized prior to database lock. Any deviations from the SAP after database lock will be documented in the final Clinical Study Report (CSR).

2 STUDY OVERVIEW

2.1 Study Objectives

2.1.1 Primary Objective

- To evaluate the change in hepatic fat fraction from Baseline to Week 12 based on magnetic resonance imaging-proton density fat fraction (MRI-PDFF) measurements in LPCN 1144 (Oral Testosterone (T)) treated subjects.

2.1.2 Secondary Objectives

- To evaluate whether 36 weeks of treatment with LPCN 1144 lowers nonalcoholic steatohepatitis (NASH) activity as determined from hepatic histology in adult men with NASH.
- To evaluate the change in hepatic fat fraction from baseline to Week 36 based on MRI-PDFF measurements in LPCN 1144 treated subjects

2.1.3 Safety Objectives

- To assess tolerability of oral LPCN 1144 in the study population.

2.2 Study Design

2.2.1 Overview

This is a proof of concept, Phase 2, randomized, double-blind, placebo controlled, three arm study in adult men with biopsy confirmed NASH. The study will be conducted across multiple centers in the United States.

Subjects will be randomized in 1:1:1 ratio to receive one of the following treatments:

- Treatment A: Oral LPCN 1144 total daily dose of 450 mg testosterone undecanoate (TU) administered as BID
- Treatment B: Oral LPCN 1144 d-alpha tocopherol formulation, total daily dose of 450 mg TU administered as BID
- Treatment C: Oral matching placebo administered as BID

There are four phases of study:

- Screening Phase for Subject requiring testosterone/androgen washout up to 12 weeks (12 weeks following intramuscular androgen injections; 4 weeks following topical or buccal androgens; 3 weeks following oral androgens).
- Screening Phase for eligibility assessments (8 weeks prior to randomization)

- Randomization Phase (one visit)
- Treatment Phase (eight visits)
- Follow-up (one visit)

Subjects will undergo a screening period to determine study eligibility. As a part of screening, liver biopsies will be performed for subjects who have not had a liver biopsy within 4 months of Screening Visit 1. Adult male subjects with histologic evidence of NASH will be enrolled into the study. Baseline fat fraction will be measured by MRI-PDFF in eligible subjects. If a subject is unable to undergo MRI-PDFF or DXA imaging due to health condition (e.g. claustrophobia, difficulty holding breath during the procedure, etc.), the missing MRI-PDFF will be treated as missing data, and the subject will be allowed to proceed to randomization or continue therapy upon investigator and medical monitor approval.

Eligible subjects will be randomized to one of the three treatment arms. The treatment phase will be for a duration of 36-weeks (not including screening and follow up periods) with assessments of liver biopsies, hepatic fat fraction, liver enzymes, lipid levels and other safety parameters. Safety and tolerability will be assessed throughout the study. A follow-up visit will be performed to assess adverse events, complete patient reported questionnaires including Patient Global Impression of Change (PGI-C), collect a urinalysis sample, and to obtain blood samples for clinical laboratory tests including, lipid panel, hematology panel, metabolic and renal function panel, hepatic panel, testosterone, and hormones and related markers.

The procedures of each visit are provided in Schedule of Activities (SOA) in the section 2.3 of the protocol.

2.2.2 Randomization and Blinding

All subjects meeting the inclusion criteria and not meeting any of the exclusion criteria will be enrolled into the study. Subjects meeting the enrollment criteria will be randomly assigned to one of the three treatment arms in a 1:1:1 ratio. The randomization will be carried out by central assignment. The study is a blinded study; therefore, all the randomization codes will be centrally maintained and no data from the randomization will be available to Sponsor, contract research organization (CRO) operations team, medical monitors, monitors or any site staff. The treatment assigned to the study subject will not be available to anyone except pre-identified study personnel.

All randomization codes will be centrally assigned and stored. Each subject will receive a unique treatment ID that does not directly associate with the treatment assigned.

2.2.3 Study Drug

LPCN 1144 is an oral capsule formulation of TU administered as 225 mg TU two times a day for a total daily dose of 450 mg TU. Two different formulations of LPCN 1144 will be evaluated, along with placebo.

All subjects should be instructed to take Treatment A (225 mg TU per dose), B (225 mg TU per dose), or C (placebo) twice daily, approximately 12 hours apart, approximately 30 minutes after morning and evening meals, with water. If a dose is missed, the subjects should take it as soon as they remember (within 4 hours of missed dose). If second dose is not taken within 16 hours of the first dose of the day, the subject should skip that dose and resume dosing at 24 hours

past the last dose. If a subject is switched to a reduced dose, the subject should take the dose in the morning, approximately 30 minutes after the morning meal, with water.

2.2.4 Sample Size Determination

This is a proof of concept Phase 2 study aimed at evaluating the LPCN 1144 as potential therapeutic agent in the treatment and / or management of NASH. The total planned sample size was approximately 75 subjects. The sample size for the study was not based on any statistical considerations but based on similar Phase 2 studies in this therapeutic area.

Based on the recruitment status, the initial target of 75 randomized subjects can not be achieved. Therefore, Lipocine decided to end randomization by September 15th 2020. Based on screening data to-date, approximately 55 subjects will be randomized in the study. With 55 subjects, the study will still be sufficiently powered (i.e, at least 80% power for each active arm vs. placebo and at least 90% power for pooled active arms vs. placebo) to detect a treatment difference of 5% in the primary outcome measure (change in MRI-PDFF from baseline at Week 12) assuming a common standard deviation of 5% at a 2-sided 0.05 significance level.

2.3 Study Outcome Measures

2.3.1 Primary Outcome Measures

- Change in hepatic fat fraction based on MRI-PDFF measurements in LPCN 1144 treated subjects after 12 weeks of treatment.

2.3.2 Key Secondary Outcome Measures

- Change in nonalcoholic fatty liver disease (NAFLD) activity evaluated via a standardized scoring of liver biopsies at baseline and after 36 weeks of treatment in LPCN 1144 treated subjects.
- Change in fibrosis evaluated via NASH clinical research network (CRN) fibrosis score of liver biopsies at baseline and after 36 weeks of treatment in LPCN 1144 treated subjects.

2.3.3 Other Secondary Outcome Measures

- Relative change in hepatic fat fraction based on MRI-PDFF measurements in LPCN 1144 treated subjects after 12 weeks of treatment.
- Change in hepatic fat fraction based on MRI-PDFF measurements in LPCN 1144 treated subjects after 36 weeks of treatment.
- Relative change in hepatic fat fraction based on MRI-PDFF measurements in LPCN 1144 treated subjects after 36 weeks of treatment.
- NAFLD resolution of subjects who at baseline is at least 5% of hepatic fat fraction with a decrease to less than 5% at end of study.
- Resolution of NASH on overall histopathological reading.
- Change in individual NAS component scores (inflammation, ballooning, steatosis) in LPCN 1144 treated subjects after 36 weeks of treatment.
- Resolution of inflammatory liver on overall histopathological reading.
- Improvement in liver fibrosis greater than or equal to one stage (NASH CRN fibrosis score).

- Change and relative change in anthropometric measurements (weight, body mass index (BMI), waist to hip ratio, waist circumference, triceps skin fold thickness, upper arm circumference) in LPCN 1144 treated subjects.
- Change and relative change in insulin resistance (assessed by Homeostasis Model Assessment (HOMA)) and Hemoglobin A1c (HbA1c) in LPCN 1144 treated subjects.
- Changes and relative changes in liver enzymes aspartate transaminase (AST), alanine transaminase (ALT), alkaline phosphatase (ALP), gamma-glutamyl transferase (GGT), Total Bilirubin (TB) and Creatine Kinase (CK) in LPCN 1144 treated subjects.
- Changes and relative changes in non-invasive markers of fibrosis and steatosis including cytokines, leptin, fibrosis markers, and lipid profile in LPCN 1144 treated subjects.
- Changes and relative changes in lipid parameters (triglycerides, very low-density lipoprotein cholesterol (VLDL-C), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), non-HDL cholesterol and free fatty acids) in LPCN 1144 treated subjects.
- Changes in Functional Activity, Health-Related Quality of Life (HR-QOL) as measured by Short Form-36 (SF-36), Patient Global Impression of Severity (PGI-S), PGI-C, and Chronic Liver Disease Questionnaire (CLDQ) and Sexual desire and distress questionnaires in LPCN 1144 treated subjects.
- Changes in laboratory parameters (clinical chemistry, hematology and urinalysis) in LPCN 1144 treated subjects.
- Subjects achieving relative reduction ≥ 30% hepatic fat fraction based on MRI-PDFF measurements.
- Subjects achieving relative reduction ≥ 23% hepatic fat fraction based on MRI-PDFF measurements.
- Change in fibrosis status evaluated via the paired biopsies after 36 weeks of treatment in LPCN 1144 treated subjects.
- Change in FibroNest scores evaluated indepently by PharmaNest after 36 weeks of treatment in LPCN 1144 treated subjects if data available.
- Changes in DXA scan parameters in LPCN 1144 treated subjects.

3 STATISTICAL METHODOLOGY

3.1 General Considerations

3.1.1 Analysis Day

Analysis day will be calculated from the date of first dose of study drug. The day of the first dose of study drug will be Day 1, and the day immediately before Day 1 will be Day -1. There will be no Day 0.

3.1.2 Analysis Visits

For efficacy parameters with limited scheduled visits (FibroScan, liver biopsy, and ELF and its components), the analysis visits will be assigned according to the scheduled visit. Note that for

subjects who early terminate, liver biopsy will only be obtained and used in analyses if the withdrawal is after Week 16, and within 56 days (or 8 weeks) after the last dose of study drug. For FibroScan, and ELF and its components, if the early term collection is >56 days (or 8 weeks) after the last dose of study drug, it will not be used in analyses.

For MRI-PDFF, dual-energy absorptiometry (DXA) scan and other efficacy parameters, scheduled visits will be assigned to analysis visits as recorded on the CRF. Unscheduled and early termination visits will be assigned to analysis visits according to the following visit windows:

Analysis Visit	Target Analysis Day	Low Analysis Day	High Analysis Day
Week 4	29	2	43
Week 8	57	44	71
Week 12	85	72	99
Week 16	113	100	127
Week 20	141	128	155
Week 24	169	156	183
Week 30	211	184	225
Week 36	253	226	259
Note: for subjects who early terminate, MRI-PDFF will only be used if the withdrawal is after Week 4, and within 56 days (or 8 weeks) after the last dose of study drug. For other efficacy parameters, if the unscheduled collection or early term collection is >56 days after the last dose of study drug, it will not be used in analyses.			

Within an analysis visit window, the measurement from the scheduled visits will be used if available. If no scheduled visit occurs, the measurement from the visit closest to the defined target analysis day within the window will be used. If there is more than one measurement with equal distance to the defined target analysis day, the latter will be used. If no visits occur within a visit window, the measurement for this visit will be treated as missing.

3.1.3 Definition of Baseline

Pre-dose measurements on Day 1 will be considered the baseline values. If the measurement at this visit is not available, the last measurement prior to the first dose of study drug may be used as the baseline value unless it is stated otherwise. The scheduled baseline visits for MRI-PDFF, FibroScan, liver biopsy, DXA scan, questionnaires, gait speed and hand grip are Screening Visits from Day -56 to -1. The scheduled baseline visits for other variables are Day 1.

3.1.4 Summary Statistics

Categorical data will generally be summarized with counts and percentages of subjects. The denominator used for the percentage calculation will be clearly defined. Continuous data will generally be summarized with descriptive statistics including n (number of non-missing values), mean, median, standard deviation, minimum, maximum.

3.1.5 Multiple Comparison/Multiplicity

For the primary outcome measure, if the difference between the mean change in MRI-PDFF from baseline at Week 12 of the pooled active treatment groups compared to the placebo group is statistically significant, a step-down procedure will be followed to test the individual

LPCN 1144 dose groups compared to each other and compared to placebo. The following testing order will be employed:

1) LPCN 1144 versus placebo
2) LPCN 1144 d-alpha tocopherol formulation versus placebo
3) LPCN 1144 d-alpha tocopherol formulation versus LPCN 1144

A two sided p-value of less than or equal to 0.05 will be considered statistically significant. No adjustments for the additional analysis or multiple comparisons will be applied for secondary outcome measure analysis; however, any significant findings must be interpreted taking into account the strength of the finding and its biologic plausibility.

3.1.6 Handling of Dropouts and Missing Data

Missing data will be reported as such. Raw data will be subjected to a statistical screen for outliers using commonly accepted procedures.

3.1.7 The Impact of Coronavirus Disease 2019 (COVID-19)

COVID-19 pandemic may impact the conduct of the study from different aspects including quarantines, site closures, travel limitations, interruptions to the supply chain for the investigational product, or other considerations if site personnel or trial subjects become infected with COVID-19.

3.2 Analysis Populations

3.2.1 Intent-to-Treat (ITT) Analysis Dataset

The ITT Analysis Dataset is defined as all randomized subjects. The ITT Analysis Dataset will be the primary analysis dataset for the analysis of efficacy data.

3.2.2 Modified Intent-to-Treat (mITT) Analysis Dataset

The mITT Analysis Dataset is defined as all randomized subjects who took at least one dose of study drug and have at least 4 weeks of on-treatment data with non-missing baseline and Week 12 MRI-PDFF.

3.2.3 Per-Protocol (PP) Analysis Datasets

The 12-Week PP Analysis Dataset is defined as subjects in the ITT Analysis Dataset who took at least one dose of study drug and complied with the protocol sufficiently to ensure that the data is likely to represent the effects of study intervention. The following criteria will be evaluated for compliance to the protocol prior to unblinding of the treatment allocation:

- Valid hepatic fat fraction measured by MRI-PDFF at Baseline and Week 12;
- No major eligibility criterion violations;
- Did not discontinue study drug prior to Week 4;
- Subjects compliance to study drug during 12-week treatment period;
- Not taken any prohibited concomitant medication during 12-week treatment period;
- No other substantial protocol violations during 12-week Treatment Period.

The 36-Week PP Analysis Dataset is defined as subjects in the ITT Analysis Dataset who took at least one dose of study drug and complied with the protocol sufficiently to ensure that the data is likely to represent the effects of study intervention. The following criteria will be evaluated for compliance to the protocol prior to unblinding of the treatment allocation:

- Valid hepatic fat fraction measured by MRI-PDFF at Baseline and Week 36;
- No major eligibility criterion violations;
- Did not discontinue study drug prior to Week 28;
- Subjects who are <80% or >120% compliant with study drug during the treatment period;
- Not taken any prohibited concomitant medication during the treatment period;
- No other substantial protocol violations during the treatment period.

The 36-Week PP Analysis Dataset - Biopsy will be defined with similar criteria with valid liver biopsy at Baseline and Week 36.

A list of subjects with major protocol deviations leading to exclusion from the 12-Week PP Analysis Dataset, the 36-Week PP Analysis Dataset and the 36-Week PP Analysis Dataset - Biopsy will be finalized prior to unblinding the randomized treatment assignments.

3.2.4 Safety Analysis Dataset

The Safety Analysis Dataset is defined as all subjects who took at least one dose of study drug. All safety data will be analyzed using the Safety Analysis Dataset. In the event that a subject takes the wrong study drug (i.e., did not take the randomized study drug), the actual treatment received will be used for analysis.

3.3 Subject Data and Study Conduct

3.3.1 Subject Disposition

Counts and percentages of subjects who were screened (signed informed consent), participated in the optional Washout, discontinued early during screening (screen failures), and randomized will be summarized in total based on all screened subjects. Reasons for early discontinuation will also be summarized. In addition, discontinuation due to COVID-19 impact will be summarized.

Counts and percentages of subjects who were randomized, started study drug, discontinued early from the study, completed Week 12, and completed the study will be summarized by treatment, pooled active treatment (two formulations of LPCN 1144), and in total based on all randomized subjects. Reasons for early discontinuation will also be summarized. In addition, early discontinuation from the study due to COVID-19 impact will be summarized.

3.3.2 Protocol Deviations

Protocol deviations will be defined in the Protocol Deviation Plan. Counts and percentages of subjects with CSR reportable protocol deviations by deviation category will be summarized by treatment, pooled active treatment (two formulations of LPCN 1144), and in total based on all randomized subjects. All protocol deviations and COVID-19 related protocol deviations will be listed.

3.3.3 Analysis Populations

Counts and percentages of subjects in each analysis set will be summarized by treatment, pooled active treatment (two formulations of LPCN 1144), and in total based on all randomized subjects.

3.3.4 Demographic and Baseline Characteristics

The following demographic and baseline characteristics will be summarized:

- Age (years) and age categories (<65 years, ≥65 years)
- Sex
- Race
- Ethnicity
- Height (cm)
- Weight (kg)
- BMI (kg/m^2) and BMI categories (<30 kg/m^2, ≥30 kg/m^2)
- Waist circumference (cm)
- Hip circumference (cm)
- Waist-to-hip ratio
- Diabetes mellitus Type 1 and Type 2 diagnosis
- Hypertension diagnosis
- Hepatic fat fraction (%) by MRI-PDFF and MRI-PDFF categories (≥5%, ≥8%, and ≥10%)
- Liver Biopsy Inclusion Criteria (Stage 1 fibrosis with NAS ≥ 4 with at least 1 point each in inflammation and ballooning, Stage 2-3 fibrosis with at least 1 point in inflammation)
- Fibrosis stage
- NAS and NAS components
- Liver fibrosis (kPa)
- CAP (dB/m)
- Total testosterone (ng/dL)
- Hemoglobin A1c (%)
- ALT (U/L)
- AST (U/L)

Demographic and baseline characteristics will be summarized with descriptive statistics or counts and percentages of subjects as appropriate by treatment, pooled active treatment (two formulations of LPCN 1144), and in total for all randomized subjects and each defined analysis population.

3.3.5 Medical History

Medical history will be coded to system organ class and preferred term using the Medical Dictionary for Regulatory Activities (MedDRA) version 22.0. Counts and percentages of subjects with medical history by system organ class and preferred term will be summarized by treatment, pooled active treatment (two formulations of LPCN 1144), and in total based on all randomized subjects.

3.3.6 Concomitant Medications

Concomitant medications will be coded to anatomical therapeutic chemical (ATC) class and preferred term using the World Health Organization (WHO) Drug Dictionary version 2019G B3. For summary purposes, the medication will be reported in both the prior and concomitant medications if the end date of a prior medication occurs after treatment starts; otherwise, medications will be considered prior medications if they stopped prior to the first dose of study drug and concomitant medications if they were taken at any time after the first dose of study drug (i.e. started prior to the first dose of study drug and were ongoing or started after the first dose of study drug).

Counts and percentages of subjects taking prior and concomitant medications by ATC class and preferred term will be summarized separately by treatment, pooled active treatment (two formulations of LPCN 1144), and in total based on the Safety Analysis Dataset. Furthermore, counts and percentages of subjects with changes in baseline dose of blood pressure medication or introduction of new blood pressure medication by ATC class and preferred term will be summarized separately by treatment, pooled active treatment (two formulations of LPCN 1144), and in total based on the Safety Analysis Dataset. The use of any prior medication or concomitant medication will be listed for all randomized subjects.

3.3.7 Study Drug Exposure and Compliance

Days of exposure to study drug will be calculated as:

date of very last dose of study drug - date of very first dose of study drug + 1

Note that the exposure calculation is intended to describe the length of time a subject was exposed to study drug and therefore does not take study drug interruptions into account. Days of exposure to study drug will be summarized by treatment based on the Safety Analysis Dataset with descriptive statistics and with counts and percentages of subjects with exposure in the following categories:

- >0 to ≤4 weeks (1 – 28 days)
- >4 to ≤8 weeks (29 – 56 days)
- >8 to ≤12 weeks (57 – 84 days)
- >12 to ≤16 weeks (85 – 112 days)
- >16 to ≤20 weeks (113 – 140 days)
- >20 to ≤24 weeks (141 – 168 days)
- >24 to ≤30 weeks (169 – 210 days)
- >30 to ≤36 weeks (211 – 252 days)
- >36 weeks (>252 days)

The total days of actual dosing and days of interrupted therapy will also be captured.

Days of actual dosing will be calculated as:

date of very last dose of study drug - date of very first dose of study drug + 1 - days of interruption

Days of interruption will be calculated as:

Sum of (date of restarting dosing of study drug - date of last dose of study drug before stopping - 1)

Percent compliance to the study drug regimen will be calculated as

100 x number of actual capsules taken / number of expected capsules taken

Where the number of actual capsules taken = number of capsules dispensed – number of capsules returned, and the number of expected capsules taken = (date of last dose – date of first dose + 1) x 2 capsules per day.

If study drug is not returned, the number of capsules returned will be considered 0 for the compliance calculation. If subject has dose reduced to 1 capsule per day, the number of expected capsules taken = (last dose date of 2 capsules per day – date of first dose + 1) x 2 capsules per day + (date of last dose – start date of 1 capsule per day + 1) x 1 capsule per day.

The percent compliance to the study drug regimen during the 12-week treatment period and the 36-week treatment period will be summarized by treatment, and pooled active treatment (two formulations of LPCN 1144) based on the Safety Analysis Dataset with descriptive statistics and with counts and percentages of subjects with compliance in the following categories:

- <80%
- 80-120%
- >120%

3.4 Efficacy Assessment

Efficacy data will be summarized and analyzed by randomized treatment, and pooled active treatment (two formulations of LPCN 1144) based on the ITT Analysis Dataset unless otherwise specified.

3.4.1 Primary Efficacy Outcome Measures

The primary efficacy outcome measure will be the change in hepatic fat fraction based on MRI-PDFF measurements from baseline to Week 12.

Change in fat fraction by MRI-PDFF = Week 12 - Baseline

Relative (percent) change in fat fraction by MRI-PDFF = 100 x (Week 12 - Baseline) / Baseline

MRI-PDFF imaging will be performed at Screening, Week 12, and Week 36/Early Terminate (ET). Summary statistics at all scheduled visits for the values, change from baseline and relative change from baseline will be provided. The summary will also be repeated based on the mITT and PP Analysis Datasets. Mean and mean changes will be plotted over time. Waterfall plots will be provided for relative changes at Week 12 and Week 36.

3.4.1.1 Primary Analysis

The primary efficacy outcome measure will be analyzed using an analysis of covariance (ANCOVA) model with treatment group as a factor and baseline hepatic fat fraction (%) based on MRI-PDFF measurements as a covariate. The 2 active arms (two formulations of LPCN 1144) will be pooled and compared to placebo. The treatment effect estimate, 95% CI and p

values will be provided. The two formulations of LPCN 1144 will also be compared to each other and compared to placebo independently using linear contrasts. Multiple imputation will be performed if MRI-PDFF is missing at baseline and/or Week 12 in the ITT Analysis Dataset. The SAS sample code is listed:

```
*********************************************************************
  TREATMENT: 0 (Placebo), 1 (LPCN 1144 450 mg), or
             2 (LPCN 1144 450 mg, d-alpha tocopherol formulation)
  BASE: Baseline fat fraction by MRI-PDFF
  CHG: Change from baseline to Week 12 in fat fraction by MRI-PDFF
*********************************************************************;
proc mixed;
  class TREATMENT;
  model CHG = TREATMENT BASE;
  lsmeans TREATMENT / cl alpha = 0.05;
  estimate "LPCN 1144 450 mg: Placebo" TREATMENT -1 1 0 /cl;
  estimate "LPCN 1144 450 mg d-alpha : Placebo" TREATMENT -1 0 1 /cl;
  estimate "LPCN 1144 450 mg d-alpha : LPCN 1144 450 mg" TREATMENT 0 -1 1
/cl;
run;
```

The Shapiro-Wilk normality test will be performed for the residuals from the model above. If the p-value <0.01, non-parametric analyses using Wilcoxon rank sum test may be explored.

Multiple Imputation (MI) Method:

Missing primary efficacy outcome measure will be handled using the Missing at random (MAR) multiple imputation method.

The multiple imputation procedure will be conducted by two parts: PROC MI and PROC MIANALYZE.

- Multiple imputation will be used to impute missing data and will be implemented using SAS® PROC MI: Missing data may be the result of missing hepatic fat fraction (%) assessed by MRI-PDFF at any time point including baseline and/or week 12, or the result of patients discontinuing treatment prior to Week 12. MAR will be assumed. The imputation model will include the subject demographics, and baseline and Week 12 hepatic fat fraction assessed by MRI-PDFF. Missing data will be imputed 100 times to generate 100 complete data sets. The seed will be 98765. The fully conditional specification method will be used. The 100 complete data sets will then be analyzed using the same analysis method as the one used to analyze the primary outcome measure.

  The SAS sample code is listed:

```
*********************************************************************
  TREATMENT: 0 (Placebo), 1 (LPCN 1144 450 mg), or
             2 (LPCN 1144 450 mg, d-alpha tocopherol formulation)
  BASE: Baseline fat fraction by MRI-PDFF
  WEEK12: Week 12 fat fraction by MRI-PDFF
  AGEGR: 1 (< Median Age), 2 (>= Median Age)
  RACEGR: 1 (White), 2 (Non-white)
  WGTBLGR: 1 (< Median Weight), 2 (>= Median Weight)
*********************************************************************;
```

```
proc mi seed=98765 nimpute=100 minimum=(. . . . 0 0) maximum=(. . . . 100
100);
  class TREATMENT AGEGR RACEGR WGTBLGR;
  FCS reg(BASE WEEK12);
  var TREATMENT AGEGR RACEGR WGTBLGR BASE WEEK12;
run;
```

- The results from the 100 fitted models will be combined using SAS® PROC MIANALYZE: The estimate of the treatment difference and standard error from the analysis will be analyzed by PROC MIANALYZE to obtain the overall estimate of treatment difference, as well as the confidence interval and p-value for the hypothesis testing.

  The SAS sample code is listed:

```
*********************************************************************
  TREATMENT: 0 (Placebo), 1 (LPCN 1144 450 m), or
             2 (LPCN 1144 450 mg, d-alpha tocopherol formulation)
  ESTIMATE: Estimated reatment difference from 100 fitted models
  STDERR: Standard error of the estimate
*********************************************************************;
proc mianalyze;
    by TREATMENT;
    modeleffects ESTIMATE;
    stderr STDERR;
run;
```

If the model fails to converge, exclusion of subject demographics will be considered.

3.4.1.2 Sensitivity Analyses

The primary analysis will be repeated on observed data using the mITT and 12-Week PP Analysis Datasets. No imputation will be performed.

Tipping-point analysis approach may be used to explore the robustness of the primary analysis results using the multiple imputation method. The assumptions in this sensitivity analysis are “missing not at random” for dropouts in the active treatment group and “missing at random” for dropouts in the placebo group. A penalty of delta ranging from 0% to 40% (or higher if necessary) by intervals of 2%, will be added to the imputed values at Week 12 in each active treatment group while the fix penalty delta = 0% will be used in the placebo group (delta = 0% corresponds to the missing at random assumption).

3.4.2 Key Secondary Efficacy Outcome Measures

The key secondary efficacy outcome measures are

- Change in NAFLD activity evaluated via a standardized scoring of liver biopsies from baseline to Week 36
- Change in fibrosis evaluated via NASH CRN fibrosis score of liver biopsies at baseline and after 36 weeks of treatment in LPCN 1144 treated subjects.

The NAS ranges from 0 to 8 (highest activity) and is calculated as the sum of three components of the standardized histologic feature scoring system for liver biopsies:

NAS = Steatosis score (0-3) +

Lobular inflammation score (0-3) +

Hepatocyte ballooning score (0-2)

Fibrosis score evaluated via NASH CRN includes 5 stages: F0, F1 (F1a, F1b, F1c), F2, F3 and F4.

Biopsy will be performed at Screening and Week 36/ET. Values and changes from baseline in NAS and fibrosis will be summarized descriptively, as well as the counts and percentages of subjects by treatment group, and pooled active treatment (two formulations of LPCN 1144).

The analysis will include both improvement in NAS and fibrosis, and numerical change in NAS and fibrosis based on the ITT. The missing value at Week 36 in ITT will be imputed by last-observed-carried-forward (LOCF) method. The analysis will be repeated on the 36-Week PP Analysis Dataset – Biopsy, and no imputation will be performed.

3.4.2.1 Improvement in NAS and Fibrosis

The definition of improvement in NAS after treatment as the key secondary outcome measure requires the following three conditions:

- No worsening of the fibrosis feature score.
- Improvement by at least 1 point in the hepatocyte ballooning feature score.
- Either:
    1. Improvement in NAS by 2 or more points spread across at least two of the NAS components, OR
    2. Post-treatment NAS is 3 points or less.

Similarly, improvement in fibrosis will be defined as improvement in liver fibrosis greater than or equal to one stage using the NASH CRN fibrosis score with no worsening of ballooning, inflammation, or steatosis.

The analysis for improvement of NAS will only include subjects with a baseline hepatocyte ballooning score of 1 or 2. Improvement in NAS and fibrosis from baseline to Week 36/LOCF will be analyzed as a binary outcome (improved vs. not improved). Counts and percentages of subjects will be presented by treatment group. The binary outcomes will be compared using Fisher's exact test. The SAS sample code is listed:

```
*****************************************************************
  TREATMENT: 0 (Placebo), 1 (LPCN 1144 450 mg), or
             2 (LPCN 1144 450 mg, d-alpha tocopherol formulation)
  RESPONDER: 0 (Non-Improvement), 1 (Improvement)
*****************************************************************;
proc freq;
  tables TREATMENT*RESPONDER / fisher;
run;
```

If concerns about confounding arise,, a logistic regression model with treatment group as a factor and baseline value as a covariate will be used. The odds ratio, 95% confidence interval, and p-value from the logistic regression will be provided. Missing value will be imputed as not improved for both methods. The SAS sample code is listed:

```
*********************************************************************
  TREATMENT: 0 (Placebo), 1 (LPCN 1144 450 mg), or
             2 (LPCN 1144 450 mg, d-alpha tocopherol formulation)
  BASE: Baseline NAS/fibrosis
  RESPONSE: 0 (Non-Improvement), 1 (Improvement)
*********************************************************************;
proc logistic;
  class TREATMENT;
  model RESPONSE (event='1') = TREATMENT BASE / link=logit
        clodds=wald orpvalue;
  oddsratio TREATMENT /diff=ref;
run;
```

3.4.2.2 Numerical Change in NAS and Fibrosis

Mean changes from baseline to Week 36/LOCF in NAS and fibrosis will be analyzed by an ANCOVA model with treatment group as a factor and baseline value as a covariate. LS means vs placebo, 95% confidence intervals and p-values will be presented.

Addtionally, median changes in NAS and fibrosis will be analyzed by Wilcoxon rank sum test with Hodges-Lehmann method. Median difference, 95% confidence intervals and p-values will be presented. No imputation will be performed. The SAS sample code is listed:

```
*********************************************************************
  TREATMENT: 0 (Placebo), 1 (LPCN 1144 450 mg), or
             2 (LPCN 1144 450 mg, d-alpha tocopherol formulation)
  CHG: Change from baseline to Week 36 in NAS/fibrosis
*********************************************************************;
proc npar1way alpha=.05 wilcoxon hl;
  class TREATMENT;
  var CHG;
run;
```

3.4.3 Other Secondary Efficacy Outcome Measures

3.4.3.1 Change and Relative Change in Hepatic Fat Fraction based on MRI-PDFF

- Relative change in hepatic fat fraction based on MRI-PDFF measurements in LPCN 1144 treated subjects after 12 weeks of treatment.
- Change in hepatic fat fraction based on MRI-PDFF measurements in LPCN 1144 treated subjects after 36 weeks of treatment.
- Relative change in hepatic fat fraction based on MRI-PDFF measurements in LPCN 1144 treated subjects after 36 weeks of treatment.

The three outcome measures are the change from baseline at Week 36 and the relative change from baseline at Week 12 or 36, which will be analyzed using ANCOVA model in a manner similar to the primary analysis of the primary outcome measure.

A mixed-model repeated-measures (MMRM) will be used for analyzing the change and relative change from baseline to Week 12 and 36 for the ITT analysis population and repeated on the mITT and PP analysis populations. No imputation will be performed. The factors in the model will be treatment group, visit, and the treatment group by visit interactions, and the covariate will be baseline value. An unstructured covariance matrix will be used (TYPE=UN). The sample SAS code can be found below:

```
**********************************************************************
  USUBJID: Unique subject identifier
  TREATMENT: 0 (Placebo), 1 (LPCN 1144 450 mg), or
             2 (LPCN 1144 450 mg, d-alpha tocopherol formulation)
  BASE: Baseline fat fraction by MRI-PDFF
  VISIT: Week 12, Week 36
  CHG: Change from baseline to Week 36 in fat fraction by MRI-PDFF
**********************************************************************;
proc mixed;
  class USUBJID TREATMENT VISIT;
  model CHG = TREATMENT BASE VISIT TREATMENT*VISIT;
  Repeated VISIT / TYPE=UN sub=USUBJID;
  lsmeans VISIT*TREATMENT / cl pdiff slice=VISIT;
run;
```

3.4.3.2 NAFLD Resolution

- NAFLD resolution of subjects who at baseline is at least 5% with a decrease to less than 5% at end of study.

Resolution of NAFLD at Week 12 or 36 is defined as MRI-PDFF at baseline ≥ 5% and MRI-PDFF at Week 12 or 36 < 5%. Counts and percentages of subjects will be presented by treatment group, and pooled active treatment (two formulations of LPCN 1144). The logistic regression model with treatment group as a factor and baseline MRI-PDFF value as a covariate will be used. The odds ratio, 95% confidence interval, and p-value will be provided. Missing value will be imputed as non-resolution.

3.4.3.3 NASH Resolution

Resolution of NASH on overall histopathological reading is defined as

- Resolution of steatohepatitis on overall histopathological reading (defined as NAS score of 0-1 for inflammation, 0 for ballooning, and any value for steatosis) and no worsening of liver fibrosis on NASH CRN fibrosis score.

OR

- Improvement in liver fibrosis greater than or equal to one stage (NASH CRN fibrosis score) and no worsening of steatohepatitis (defined as no increase in NAS for ballooning, inflammation, or steatosis)

The analysis for resolution of NASH will only include subjects with a baseline NAS ≥ 4 with lobular inflammation score ≥1 and hepatocyte ballooning score ≥1 and any value for steatosis score.

Counts and percentages of subjects will be presented by treatment group, and pooled active treatment (two formulations of LPCN 1144). Fisher's exact test will be used for the analysis at Week 36/LOCF. Odds ratios (vs Placebo), 95% confidence intervals and p-values will be provided. Missing value will be imputed as non-resolution.

3.4.3.4 Inflammatory Liver Resolution

Resolution of inflammatory liver on overall histopathological reading is defined as

- NAS score of 0-1 for inflammation, 0 for ballooning, and any value for steatosis at Week 36/ET and no worsening of fibrosis.

OR

- Improvement in liver fibrosis greater than or equal to one stage and no increase in NAS for ballooning, inflammation, or steatosis

The analysis for inflammatory liver resolution will only include subjects with a baseline lobular inflammation score >1 and baseline hepatocyte ballooning score >0 and any value for steatosis score regardless of total NAS score.

Counts and percentages of subjects will be presented by treatment group, and pooled active treatment (two formulations of LPCN 1144). Fisher's exact test will be used for the analysis at Week 36/LOCF. Odds ratios (vs Placebo), 95% confidence intervals and p-values will be provided. Missing value will be imputed as non-resolution.

3.4.3.5 NAS component scores

- Change in individual NAS component scores (inflammation, ballooning, steatosis) in LPCN 1144 treated subjects after 36 weeks of treatment.

Biopsy will be performed at Screening and Week 36/ET. Values and changes from baseline in individual NAS component scores will be summarized descriptively. The changes in individual NAS component scores will be analyzed by an ANCOVA model with treatment group as a factor and baseline value as a covariate. LS means vs placebo, 95% confidence intervals and p-values will be presented. The missing value will be imputed by LOCF method.

3.4.3.6 Liver Fibrosis Stage

- Improvement in liver fibrosis greater than or equal to one stage (NASH CRN fibrosis score).

Counts and percentages of subjects will be presented by treatment group, and pooled active treatment (two formulations of LPCN 1144). Fisher's exact test will be used for the analysis at Week 36/LOCF. The odds ratio (vs Placebo), 95% confidence interval, and p-value will be provided. Missing value will be imputed as non-improved.

3.4.3.7 Liver Enzymes

Liver enzymes (AST, ALT, ALP, GGT, TB and CK) will be collected at each visit.

Values for baseline and for values and changes from baseline will be presented for each scheduled visit. Mean and mean changes will be plotted over time. The change and relative change from baseline to Weeks 4, 8, 12, 16, 20, 24, 30 and 36 will be analyzed using a MMRM. No imputation will be performed. The factors in the model will be treatment group, visit, and the treatment group by visit interactions, and the covariate will be baseline value. An unstructured covariance matrix will be used (TYPE=UN).

3.4.3.8 Subjects achieving absolute reduction ≥ 17 U/L in ALT at Week 12 with ALT > ULN at baseline.

Counts and percentages of subjects will be presented by treatment group, and pooled active treatment (two formulations of LPCN 1144) at Week 12. Fisher's exact test will be used for the analysis at Week 12. The odds ratio (vs Placebo), 95% confidence interval, and p-value will be provided. The analysis will be performed on the observed data without imputation.

3.4.3.9 Lipid Profile

Lipid profile (total cholesterol, triglycerides, VLDL-C, HDL-C, LDL-C, and free fatty acids) will be collected at Day 1, Week 4, Week 8, Week 12, Week 16, Week 20, Week 24, Week 30, Week 36 and Week 38.

Non-HDL cholesterol will be calculated as Total Cholesterol - HDL-C.

Values for baseline and for values and changes from baseline will be presented for each scheduled visit. The change and relative change from baseline to Weeks 4, 8, 12, 16, 20, 24, 30 and 36 will be analyzed using a MMRM. No imputation will be performed. The factors in the model will be treatment group, visit, and the treatment group by visit interactions, and the covariate will be baseline value. An unstructured covariance matrix will be used (TYPE=UN).

For the relative change in triglycerides, the MMRM specified above will be performed on the log-transformed data, and the results will be back-transformed.

3.4.3.10 Fibrosis and Steatosis Biomarkers

Fibrosis and steatosis biomarkers including ELF [and its components Hyaluronic Acid (HA), N-terminal propeptide of type III collagen (PIIINP), Tissue inhibitor of metalloproteinase 1 (TIMP-1)], Transforming growth Factor-Beta (TGF-Beta), Tissue Necrosis Factor-alpha (TNF-alpha) will be collected at Day 1 and Week 36/ET.

Values for baseline and for values and changes from baseline will be presented for Week 36. ANCOVA model with treatment group as a factor and baseline value as a covariate will be used for analyzing the change and relative change from baseline to Week 36. LS means vs placebo, 95% confidence intervals and p-values will be presented. The missing value will be imputed by LOCF method.

3.4.3.11 Insulin Resistance (assessed by HOMA) and HbA1c

Insulin Resistance (assessed by HOMA) and HbA1c will be collected at Day 1, Week 12 and Week 36/ET.

Values for baseline and for values and changes from baseline will be presented for each scheduled visit. ANCOVA model with treatment group as a factor and baseline value as a covariate will be used for analyzing the change and relative change from baseline to Week 12, and the change and relative change from baseline to Week 36. LS means vs placebo, 95% confidence intervals and p-values will be presented. The missing value will be imputed by LOCF method. The same ANCOVA model will be repeated on the observed data without imputation.

3.4.3.12 Anthropometric Measurements

The anthropometric measurements include weight, BMI, waist to hip ratio, waist circumference, triceps skin fold thickness, upper arm circumference. The measurements of weight, and waist and hip circumference will be performed at Screening, Day 1, Week 4, Week 8, Week 12, Week 16, Week 20, Week 24, Week 30, and Week 36. The measurements of triceps skin fold thickness, and mid upper arm circumference will be performed at Screening, Week 12, and Week 36.

Values for baseline and for values and changes from baseline will be presented for each scheduled visit.

The change and relative change from baseline to Weeks 4, 8, 12, 16, 20, 24, 30 and 36 in weight, BMI, waist to hip ratio, and waist circumference will be analyzed using a MMRM. No imputation will be performed. The factors in the model will be treatment group, visit, and the treatment group by visit interactions, and the covariate will be baseline value. An unstructured covariance matrix will be used (TYPE=UN).

ANCOVA model with treatment group as a factor and baseline value as a covariate will be used for analyzing the change and relative change from baseline to Week 12, and the change and relative change from baseline to Week 36 in the triceps skin fold thickness, and upper arm circumference. LS means vs placebo, 95% confidence intervals and p-values will be presented. The missing value will be imputed by LOCF method. The same ANCOVA model will be repeated on the observed data without imputation.

3.4.3.13 Subjects achieving relative reduction ≥ 30% or 23% hepatic fat fraction based on MRI-PDFF measurements.

Counts and percentages of subjects will be presented by treatment group, and pooled active treatment (two formulations of LPCN 1144) at Week 12 and Week 36. A logistic regression model with treatment group as a factor and baseline value as a covariate will be used. The odds ratio, 95% confidence interval, and p-value from the logistic regression will be provided. Missing value will be imputed as not improved for both methods.

3.4.3.14 Fibrosis Status Evaluated via the Paired Biopsies

The following analysis will be performed if data available.

A direct comparison of a patient's two biopsies to each other at the end of study can provide a complimentary assessment of the difference between the two biopsies and allow a qualitative determination of intra grade or intrastage changes. Paired biopsies from each patient will be randomly sorted and assigned as either A or B. The pathologist will be unaware of which biopsy is which. The assessment will be made as better, worse or same for the changes in Fibrosis and the three components including Steatosis, Lobular Inflammation and Ballooning.

Counts and percentages of subjects will be presented by treatment group, and pooled active treatment (two formulations of LPCN 1144). Cochran-Mantel-Haenszel (CMH) test will be used for the analysis and the row mean scores differ p-value will be provided. The sample SAS code can be found below:

```
*****************************************************************
  TREATMENT: 0 (Placebo), 1 (LPCN 1144 450 mg), or
             2 (LPCN 1144 450 mg, d-alpha tocopherol formulation)
  RESPONSE: -1 (Worse), 0 (Same), 1 (Better),
*****************************************************************;
proc freq order=data;
  tables Treatment*Response / cmh score=modridit;
  weight count;
  ods output cmh=cmh;
run;
```

3.4.3.15 FibroNest Scores Evaluated by PharmaNest

The following analysis will be performed if data available.

Values for baseline and for values and changes from baseline will be presented for each scheduled visit. ANCOVA model with treatment group as a factor and baseline value as a covariate will be used for analyzing the change from baseline to Week 36. LS means, 95% confidence intervals and p-values will be presented. The missing value will be imputed by LOCF method.

3.4.3.16 Dual-energy X-ray Absorptiometry (DXA) Scan

DXA scan including whole body lean mass, appendicular lean mass, fat mass, and bone mineral density will be performed at Screening, Week 20 and Week 36.

DXA scan parameters will be defined/calculated as below:

DXA PARAMETERS	MICL TEST CODE	MICL TEST NAME	MICL TEST UNITS
Whole Body Lean Mass	TBC_TotLean_Mass	Total Lean Mass	g
Appendicular Lean Mass	calculated as: Appendicular Lean Muscle Mass = TBC_R_ArmLean_Mass + TBC_L_ArmLean_Mass + TBC_R_LegLean_Mass + TBC_L_LegLean_Mass		g
	TBC_R_ArmLean_Mass	Right Arm Lean Mass	g
	TBC_L_ArmLean_Mass	Left Arm Lean Mass	g
	TBC_R_LegLean_Mass	Right Leg Lean Mass	g
	TBC_L_LegLean_Mass	Left Leg Lean Mass	g
	TBC_TotFat_Mass	Total Fat Mass	g
Whole Body Fat Mass	Total_FNeck_BMD	Total Femoral Neck BMD	g/cm2
Bone Mineral Density (Femoral)	Total_Spine_BMD	Total Spine BMD	g/cm2
Bone Mineral Density (Lumbar)			

Values for baseline and for values, change and relative change from baseline will be presented at each scheduled visit by DXA scan parameters. ANCOVA model with treatment group as a factor and baseline value as a covariate will be used for analyzing the change and relative change from baseline to Week 20, and the change and relative change from baseline to Week 36. LS means vs placebo, 95% confidence intervals and p-values will be presented. The missing value will be imputed by LOCF method. The same ANCOVA model will be repeated on the observed data without imputation.

Counts and percentages of subjects, who have an increase from baseline in whole body lean mass and decrease from baseline in whole body fat mass, will be presented by treatment group, and pooled active treatment (two formulations of LPCN 1144) at Week 20 and Week 36. Fisher's exact test will be used for the analysis at Week 20 and Week 36. The odds ratio (vs Placebo), 95% confidence interval, and p-value will be provided. The analysis will be performed on the observed data without imputation.

3.4.4 Subgroups

The analyses of the primary outcome measure will be performed for the following subgroups:

- Subjects with baseline liver fat ≥ 5%;
- Subjects with baseline liver fat ≥ 8%;
- Subjects with baseline liver fat ≥ 10%;
- Baseline serum total Testosterone level (< median, ≥ median, < 300 ng/dL, ≥ 300 ng/dL);
- Diabetes Status at Screening = Type 1 or 2;

- Baseline ALT > ULN;
- Baseline BMI >=30 kg/m2; and
- Ethnicity = Hispanic or Latino.

The analyses of the primary outcome measure may be evaluated for the COVID-19 subgroup as needed.

The analyses of the key secondary outcome measures will be performed for the following subgroups:

- Baseline Inflammation >=1 and Ballooning >=1 for the improvement in NAS;
- Baseline NAS ≥ 4 for the improvement in NAS; and
- Baseline fibrosis stage (F1, F2-3) for the improvement in Fibrosis.

The analyses of the other outcome measure - Relative Change in Hepatic Fat Fraction based on MRI-PDFF will be performed for the following subgroups:

- Subjects with baseline liver fat ≥ 10%; and
- Ethnicity = Hispanic or Latino.

The analyses of the other secondary outcome measure - Liver Enzymes will be performed for the following subgroups:

- Baseline ALT > ULN

3.5 Quality of Life Assessment

Functional Activity, HR-QOL (SF-36) and CLDQ questionnaires will be administered at Screening and Week 36. PGI-S and Sexual desire and distress questionnaires will be administered at each study visit except Day 1 and Week 38. PGI-C questionnaire will be administered at Week 12 Week 36 and Week 38. I-PSS questionnaire will be administered at Screening, Week 12 and Week 36. AUDIT (interview version) and Skinner Alcohol questionnaire will be administered at Screening. AUDIT-C questionnaire will be administered at Week 12.

Quality of life data will be summarized and analyzed by randomized treatment, and pooled active treatment (two formulations of LPCN 1144) based on the ITT Analysis Dataset unless otherwise specified.

Values and changes from baseline in Functional Activity, HR-QOL (SF-36), CLDQ and I-PSS will be presented at each scheduled visit and baseline by domain or total score. Values and changes from baseline in PGI-S, PGI-C, and sexual desire and distress questionnaires will be presented at each scheduled visit and baseline by item. All questionnaires responses will be listed.

For HR-QOL (SF-36) and CLDQ questionnaires, an ANCOVA model with treatment group as a factor and baseline value as a covariate will be used for analyzing the change from baseline to Week 36. LS means vs placebo, 95% confidence intervals and p-values will be presented. No imputation will be performed.

3.5.1 Functional Activity Questionnaire (NHANES III)

The questionnaire has 19 items grouped into eight domains: Vigorous work (PAQ 605, 610, 615), Moderate work (PAQ 620, 625, 630), Travel (PAQ 635, 640, 645), Vigorous recreation (PAQ 650, 655, 660), Moderate recreation (PAQ 665, 670, 675), Sitting (PAQ 680), Physically activity (PAQ 706), and Sedentary behavior (PAQ 710, 715).

3.5.2 Health-Related Quality of Life (HR-QOL) measured by Short Form-36

The SF-36 questionnaire is to measure functional health and well-being from the subject's point of view. It consists of eight health domains in the table below. These health domain scales contribute to the physical health and mental health summary measures.

The scores are weighted sums of the questions in each domain. All the scores can range from 0 - 100. Higher scores represent better health. The Physical Component Summary (PCS) was calculated by positively weighting the 4 subscales in the physical domain (PF, RP, BP, and GH) and the remaining psychological domain subscales negatively. In contrast, the Mental Component Summary (MCS) was calculated by positively weighting the 4 mental domain subscales (VT, SF, RE, and MH) and negatively weighting the 4 physical domain subscales. The domain scores are derived using an uncorrelated factor solution. The correlated PCS and MCS are derived from an obliquely rotated factor solution and adjusted by 1998 US general population norms, age and gender.

Domains	Number of Items	Items
Physical Functioning (PF)	10	3, 4, 5, 6, 7, 8, 9, 10, 11, 12
Role-Physical (RP)	4	13, 14, 15, 16
Bodily Pain (BP)	2	21, 22
General Health (GH)	5	1, 33, 34, 35, 36
Vitality (VT)	4	23, 27, 29, 31
Social Functioning (SF)	2	20, 32
Role-Emotional (RE)	3	17, 18, 19
Mental Health (MH)	5	24, 25, 26, 28, 30

3.5.3 Chronic Liver Disease Questionnaire (CLDQ)

The questionnaire has 29 items grouped into six domains: AS, FA, SS, AC, EF and WO. Items are scored by a 7-point rating scale from 1 (All of the time) to 7 (None of the time).

Domains	Number of Items	Items
Abdominal Symptoms (AS)	3	1, 5, 17
Fatigue (FA)	5	2, 4, 8, 11, 13
Systemic Symptoms (SS)	5	3, 6, 21, 23, 27
Activity (AC)	3	7, 9, 14
Emotional Function (EF)	8	10, 12, 15, 16, 19, 20, 24, 26
Worry (WO)	5	18, 22, 25, 28, 29

3.5.4 Patient Global Impression of Severity (PGI-S)

It is a single question asking the patient to rate the overall impression of their health on a 5-point scale of 1 (None) to 5 (Very severe).

3.5.5 Patient Global Impression of Change (PGI-C)

It is a single question asking the patient to rate the overall status on a 7-point scale of 3 (Very much better) to -3 (Very much worse).

3.5.6 Sexual Desire and Distress Questionnaire

Daily Rating is a single question asking the patient to rate the level (degree) of sexual desire today on a 6-point scale of 0 (None at all) to 5 (Very high). Weekly Rating is a single question asking the patient to rate the feeling of being bothered by low sexual desire over the past 7 days on a 5-point scale of 0 (Never) to 4 (Always).

3.5.7 International Prostate Symptom Score (I-PSS)

The questionnaire has two parts: International Prostate Symptom Score with 7 items and Quality of Life Due to Urinary Symptoms with 1 item. Total IPSS score ranges from 0 to 35, which can be categorized as mild (0 - 7), moderate (8 -19) and severe (20 - 35). The quality of life assessment index L ranges from 0 (Delighted) to 6 (Terrible).

3.6 Safety Assessment

Safety data will be summarized by actual treatment received, and pooled active treatment (two formulations of LPCN 1144) based on the Safety Analysis Dataset.

3.6.1 Adverse Events (AEs)

AEs will be captured from the date of informed consent through study completion. All AEs will be coded to system organ class and preferred term using MedDRA version 22.0.

Treatment-emergent adverse events (TEAEs) are defined as AEs that start during a unique treatment or that already exist before the start of that unique treatment but worsen in severity or relationship to the unique treatment during the treatment.

Adverse events of special interest (AESIs) include cardiovascular events, drug induced liver (DILI) events and elevated hematocrit related to plasma volume depletion.

An overview of TEAEs will be provided including counts and percentages of subjects with the following:

- Any TEAEs (overall and by maximum severity)
- Any study drug related TEAEs (overall and by maximum severity)
- Any TEAEs of special interest (cardiovascular, DILI and elevated hematocrit)
- Any treatment-emergent serious AEs (TESAEs)
- Any TEAEs leading to discontinuation of study drug
- Any TEAEs leading to premature discontinuation from the study
- Any TEAEs leading to death

Counts and percentages of subjects will also be presented by system organ class and preferred term for each of the categories in the overview.

Listings will be presented specifically for all AEs, AESIs, SAEs and deaths, TESAEs and AEs leading to premature discontinuation from the study.

3.6.2 Clinical Laboratory Tests

C-Reactive Protein, fasting glucose and insulin will be collected at Day 1, Week 12 and Week 36.

Hepatic panel (TB, direct bilirubin, ALT, AST, ALP, GGT, Bile acid and CK), hematology panel (hematocrit, hemoglobin, prothrombin time (PT), international normalized ratio (INR)), metabolic and renal function panel (sodium, potassium, chloride, bicarbonate, calcium, phosphate, Blood Urea Nitrogen (BUN), creatinine, uric acid, albumin, total protein, estimated glomerular filtration rate (GFR)), testosterone assay (Total and calculated free T, SHBG), and urinalysis (Color, appearance, specific gravity, pH, protein, blood, ketones, urobilinogen, glucose, bilirubin, leukocyte esterase, nitrite) will be collected at each study visit.

Biomarkers will be collected at Day 1 and Week 36, including Monocyte Chemotactic Protein-1 (MCP-1), Enhanced Liver Fibrosis (ELF), [and components, Hyaluronic Acid (HA), N-terminal propeptide of type III collagen (PIIINP), Tissue inhibitor of metalloproteinase 1 (TIMP-1)], Transforming growth Factor-Beta (TGF-Beta), Tissue Necrosis Factor-alpha (TNF-alpha).

Lipid profile (total cholesterol, triglyceride, fatty acids, VLDL-C, LDL-C, HDL-C), and hormones and related markers (DHT, LH, FSH, TSH, SHBG) will be collected at Day 1, Week 4, Week 8, Week 12, Week 16, Week 20, Week 24, Week 30, Week 36 and Week 38.

HbA1c will be collected at Screening, Week 12 and Week 36.

PSA and prolactin will be collected at Screening, Week 4, Week 8, Week 12, Week 16, Week 20, Week 24, Week 30, Week 36 and Week 38.

Cortisol and cortisol binding globulin (CBG) will be collected at Screening, Week 4, Week 8, Week 12, Week 16, Week 20, Week 24, Week 30, and Week 36.

Drug and alcohol screen will be collected at Screening, Week 4, Week 12, and Week 24.

Serology will be collected at Screening.

Values for baseline and for values and changes from baseline will be presented for each scheduled visit. The incidence of abnormalities (as defined by normal ranges) prior to the first dose of study drug and after the first dose of study drug will be summarized with counts and percentages of subjects. A shift table will be provided for selected laboratory tests.

An evaluation of drug-induced serious hepatotoxicity (eDISH) plot will be provided to display peak serum ALT and TB levels for each subject.

3.6.3 Vital Signs

An abbreviated physical exam including measurements of height, weight, waist and hip circumference, vital signs (temperature, single measurement, respiratory rate, single measurement, clinic blood pressure, triplicate measurement, and pulse rate, triplicate measurement), examination for scleral icterus and pedal edema, and auscultation of heart and lungs will be performed at Day 1, Week 4, Week 8, Week 16, Week 20, Week 24, and Week 30. A detailed physical exam, which includes all tests in the abbreviated physical exam plus triceps skin fold thickness, mid upper arm circumference, and general physical findings (hepatosplenomegaly, peripheral manifestations of liver disease, ascites, wasting, fetor) will be performed at Screening, Week 12, and week 36. A prostate digital rectal examination will be

performed at Screening and Week 36. Triplicate measurements will be averaged prior to summary.

Values and changes from baseline will be presented at each scheduled visit and baseline by vital signs parameters.

Blood pressure will be summarized for the following the subgroups:

- Subjects who have a diagnosis of diabetes mellitus at baseline;
- Subjects who do not have a diagnosis with hypertension and are normotensive at baseline; and
- Subjects who have a diagnosis of hypertension at baseline and are controlled with medical management.

The change and relative change from baseline to Weeks 4, 8, 12, 16, 20, 24, 30 and 36 in the blood pressure will be analyzed using a MMRM. No imputation will be performed. The factors in the model will be treatment group, visit, and the treatment group by visit interactions, and the covariate in the model will be baseline value,. An unstructured covariance matrix will be used (TYPE=UN).

3.6.4 Fibroscan

Fibroscan imaging will be performed at Screening, Week 12, Week 24, and Week 36.

Values and changes from baseline of liver fibrosis will be presented at each scheduled visit and baseline.

3.6.5 Gait Speed

Gait speed will be assessed at Screening and Week 36.

Values and changes from baseline of time to complete the 10-meter walk will be presented at each scheduled visit and baseline.

3.6.6 Hand Grip Strength Test

Hand grip strength test will be assessed at Screening and Week 36.

Values and changes from baseline will be presented at each scheduled visit by hand.

3.6.7 Cardiovascular Risk Evaluation

Counts and percentages of subjects will be presented by the type of cardiovascular event.

3.6.8 Other Safety Assessments

Other safety data will be listed. Additionally, the subject visits impacted by COVID-19 will be listed.

4 DATA AND SAFETY MONITORING BOARD (DSMB)

An independent DSMB, appointed by Sponsor, is responsible for monitoring the accumulated interim data as the trial progresses to ensure subject safety and to review efficacy. The DSMB is a multidisciplinary group with a written charge provided by Sponsor. The DSMB will include three or more members, at least one of whom is a hepatologist. After the trial commences, the

DSMB will establish a charter and meet to review data, evaluate emerging safety information, or other issues.

Interim data on safety measures requested by the DSMB are reviewed at each of the scheduled meetings. Serious adverse events will be reviewed by the DSMB as they occur with the option of a teleconference discussion if any DSMB member so requests.

The DSMB will also review the overall progress of the trial in terms of recruitment and data quality and makes a formal recommendation to Sponsor at the end of each scheduled meeting as to whether the trial should continue unmodified, continue with protocol modifications, or be stopped.

5 ANALYSIS TIMING

5.1 Primary Analysis Reviews

Primary Analysis will be performed after all subjects complete the Week 12 study visit or withdraw priror to Week 12 study visit and the database has been locked for the analysis at Week 12.

An unblinded biostatistician will be involved and send the unblinded tables to the pre-specified personnel only. The unblinded individual Week 12 MRI-PDFF results will be only available for the unblinded personnel or team. The individual treatments and individual Week 12 MRI-PDFF results must remain blinded to the study team until final data base lock and full study unblinding.

The unblinded tables will be unblinded at the treatment group level and be limited to topline results for efficacy and safety (i.e. primary outcome measure, selected other secondary efficacy outcome measures like liver enzymes (ALT/AST), selected TEAE, selected exploratory endpoints). Statistics (i.e. minimum, maximum) that could potentially unblind at the individual level will be removed.

All the details will be defined in the unblinding plan.

5.2 Week 20 Analysis Reviews

Week 20 Analysis will be performed after all subjects complete the Week 20 study visit or withdraw priror to Week 20 study visit and the database has been locked for the analysis at Week 20.

An unblinded biostatistician will be involved and send the unblinded tables to the pre-specified personnel only. The individual treatments must remain blinded to the study team until final data base lock and full study unblinding.

The unblinded tables will be unblinded at the treatment group level and be limited to results for DXA efficacy. Statistics (i.e. minimum, maximum) that could potentially unblind at the individual level will be removed.

All the details will be defined in the unblinding plan.

5.3 Interim Analysis

No interim analysis is planned.

6 CHANGES FROM PROTOCOL-SPECIFIED STATISTICAL ANALYSES

Statistical methods will include pooling the 2 active arms for comparison to placebo for the primary outcome measure analysis in addition to the comparisons between each individual active arm to placebo.

Relative changes are added in the following other secondary outcome measures: anthropometric measurements, insulin resistance, non-invasive markers, and lipid parameters.

The definition of resolution of NASH is revised.

Heat-Shock Protein 47 fibroblast marker (HSP47 fibroblast marker), and Alpha-Smooth Muscle Actin (Alpha SMA) are removed from biomarkers.

The following other secondary outcome measures are added:

- Resolution of inflammatory liver on overall histopathological reading.
- Subjects achieving relative reduction ≥ 30% hepatic fat fraction based on MRI-PDFF measurements at Week 12.
- Subjects achieving relative reduction ≥ 23% hepatic fat fraction based on MRI-PDFF measurements at Week 12.
- Change in fibrosis status evaluated via the paired biopsies after 36 weeks of treatment in LPCN 1144 treated subjects.
- Change in FibroNest scores evaluated indepently by PharmaNest after 36 weeks of treatment in LPCN 1144 treated subjects if data available.
- Non-HDL cholesterol will be calculated.
- Subjects achieving absolute reduction ≥ 17 U/L in ALT at Week 12 with ALT > ULN at baseline.

The following subgroup analysis are added for the primary outcome measure:

- Diabetes Status at Screening = Type 1 or 2;
- Baseline ALT > ULN;
- Baseline BMI >=30 kg/m2; and
- Ethnicity = Hispanic or Latino.

The following subgroup analysis are added for the other outcome measure - Relative Change in Hepatic Fat Fraction based on MRI-PDFF:

- Subjects with baseline liver fat ≥ 10%; and
- Ethnicity = Hispanic or Latino.

The following subgroup analysis are added for the other secondary outcome measure - Liver Enzymes:

- Baseline ALT > ULN

DXA scan parameters are added as one ofother secondary outcome measures instead of safety. Week 20 analysis review is planned to review the DXA scan parameters.

7 PROGRAMMING SPECIFICATIONS

Analyses will be performed using SAS® version 9.4 or higher. All available data will be presented in subject data listings which will be sorted by subject and visit date as applicable. Detailed Programming Specifications will be provided in a separate document.