Statistical Analysis Plan	
Official title:	A Phase 2, Multi-Center, Double-Blind, Randomized, Placebo-Controlled Study of MGL-3196 in Patients With Non-Alcoholic Steatohepatitis
NCT Number:	NCT02912260
Document Date:	08 May 2019

STATISTICAL ANALYSIS PLAN

A Phase 2, Multi-Center, Double-Blind, Randomized, Placebo-Controlled Study of MGL-3196 in Patients With Non-Alcoholic Steatohepatitis

Investigational Product: MGL-3196

Protocol Number: MGL-3196-05

Sponsor:

Madrigal Pharmaceuticals, Inc.

Four Tower Bridge, 200 Barr Harbor Drive, Suite 400

West Conshohocken, PA 19428

Telephone: 610-220-7260

SAP Version: 1.2

SAP Date: 08 May 2019

Original Protocol: 14 July 2016

Amendment 1 Protocol: 07 Sep 2016

Amendment 2 Protocol: 16 Dec 2016

Amendment 3 Protocol: 11 May 2017

Amendment 4 Protocol: 10 August 2017

Amendment 5 Protocol: 06 November 2017

Amendment 6 Protocol: 26 December 2017

SAP SIGNATURE PAGE

STUDY TITLE: A Phase 2, Multi-Center, Double-Blind, Randomized, Placebo Controlled Study of MGL-3196 in Patients With Non-Alcoholic Steatohepatitis

We, the undersigned, have reviewed and approved this statistical analysis plan.

Signature	Date
[redacted] Statistical Analyst [redacted]	
[redacted] Chief Medical Officer, Executive Vice President R&D Madrigal Pharmaceuticals, Inc.	

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AE	Adverse event
ALT	Alanine aminotransferase
ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
Apo	Apolipoprotein
AST	Aspartate aminotransferase
ECG	Electrocardiogram
γ-GTP	Gamma-glutamyl transpeptidase
FT3	Free triiodothyronine
FT4	Free thyroxine
HbA1c	Hemoglobin A1c
HDL-C	High-density lipoprotein cholesterol
HOMA-IR	Homeostatic model assessment for insulin resistance
hsCRP	High-sensitivity C-reactive protein
LDL-C	Low-density lipoprotein cholesterol
LLN	Lower limit of normal
LOCF	Last observation carried forward
MMRM	Mixed model for repeated measurements
MRI-PDFF	Proton density fat fraction magnetic resonance imaging
T3	Triiodothyronine
T4	Thyroxine
TBG	Thyroxine-binding globulin
TC	Total cholesterol
TEAE	Treatment-emergent adverse event
TG	Triglycerides
ULN	Upper limit of normal
WHO	World Health Organization

Version History

Version	Date	Notes
1.0	May 15, 2018	SAP approved
1.1	May 21, 2018	Amendment: • Updated Nash Resolution Responder Definition by excluding the third criteria of no worsening in fibrosis and adding requirement for at least a 2-point reduction in NAS to NASH Resolution Responder. • Added endpoints related to NAS, fibrosis and NASH resolution responders.
1.2		Amendment: • Reconciled differences in SAP 1.1 in endpoints as defined in sections 2.1, 2.3 and 3.7.3 • Added exploratory analyses for Extension study

1. INTRODUCTION

This document provides a description of the statistical methods and procedures to be implemented for the analyses of data from the study with protocol number MGL-3196-05, Amendment 4, 02 August 2017. Any deviations from this Statistical Analysis Plan (SAP) will be documented in the final clinical study report.

2. STUDY OBJECTIVES AND STUDY DESIGN

2.1 Study Objectives

2.1.1 Primary Objective

The primary objective of this study is to determine the effect of once-daily oral MGL-3196 80 mg versus placebo (randomized 2:1) for 12 weeks on the percent change in hepatic fat fraction by proton density fat fraction magnetic resonance imaging (MRI-PDFF) from baseline in patients with biopsy proven NASH.

2.1.2 Secondary Objectives

The secondary objectives of the study are the following:

- To evaluate the safety and tolerability, including any changes in thyroid axis hormones, of once-daily oral MGL-3196 80 mg versus placebo after 12 and 36 weeks in patients with biopsy-proven NASH;
- To evaluate the effect of once-daily oral MGL-3196 80 mg versus placebo for 36 weeks on the percent change in hepatic fat fraction by MRI-PDFF from baseline in patients with biopsy-proven NASH;
- To evaluate the effect of once-daily oral MGL-3196 80 mg versus placebo for 12 and 36 weeks on the change in hepatic fat fraction by MRI-PDFF from baseline in patients with biopsy-proven NASH; and
- To evaluate the effect of once-daily oral MGL-3196 80 mg versus placebo for 12 and 36 weeks on the number of patients achieving >=30% relative fat reduction at Weeks 12 and 36 (MRI-PDFF Responders)
- To determine the effect of once-daily oral MGL-3196 80 mg versus placebo after 12 and 36 weeks in patients with biopsy-proven NASH on:
    - High-sensitivity C-reactive protein (hsCRP);
    - Serum alanine aminotransferase (ALT) and aspartate aminotransferase (AST);
    - Lipid parameters including low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), non-HDL-C, total cholesterol (TC), triglycerides (TG), apolipoprotein B (ApoB), and lipoprotein(a) (Lp[a]) particles, ApoCIII; and
    - NASH and fibrosis biomarkers including cytokeratin-18 (CK-18), fibrosis-4 (FIB-4), ProC3, and enhanced liver function (ELF) test.
    -

2.1.3 Liver Biopsy Secondary and Exploratory Objectives

Responder variables at Week 36 include:

- NASH Biopsy Responders:
  - The number and percentage of patients achieving a 1-point reduction in NAS;
  - the number and percentage of patients achieving a 2-point reduction in NAS at 36 weeks;
  - the number and percentage of patients achieving a 2-point reduction in NAS at 36 weeks; and either ≥1 point reduction in lobular inflammation or ≥1 point reduction in hepatocellular ballooning; and
  - the number and percentage of patients achieving a 2-point reduction in NAS at 36 weeks; and no worsening of fibrosis.
- NASH Component Responder definition (only requires one or more criteria):
  - reduction in hepatocellular steatosis; or
  - reduction in lobular inflammation; or
  - reduction in hepatocellular ballooning.
- Fibrosis Responders:
  - the number and percentage of patients achieving ≥ 1 stage reduction in fibrosis;
  - the number and percentage of patients achieving ≥ 1 stage reduction in fibrosis; and no worsening of NAS; and
  - the number and percentage of patients achieving ≥ 2 stage reduction in fibrosis.
- NASH Resolution Responder Definition (requires both criteria):
  - lobular inflammation = 0 or 1; and
  - hepatocellular ballooning = 0.
- NASH Resolution Responders with at least a 2-point reduction in NAS at Week 36.
- [redacted]
- The number and percentage of patients achieving a change in NASH diagnosis (ie, from definite NASH to not-NASH based on histology) from baseline to Week 36.

Other liver biopsy analyses include:

- [redacted]
- The change in portal inflammation from baseline at 36 weeks; and

- [redacted]
- Correlation of reduction of liver fat on MRI-PDFF at 12 and 36 weeks with improvement or resolution of NASH on liver biopsy.
- Liver biopsy analyses performed on prespecified and data driven groups.

2.1.4 Tertiary/Exploratory Objective

[redacted]

2.1.5 Extension Study Exploratory Objectives

[redacted]

2.2 Study Design

The population for this study is male and female patients ≥18 years of age with biopsy-proven NASH.

This is a multi-center, double-blind, randomized, placebo-controlled study to evaluate the safety and efficacy of MGL-3196 in patients with NASH. Patients who qualify for study inclusion will be randomized to receive one of two 36-week treatments: MGL-3196 80 mg or placebo given orally once-daily in the morning. Following 12 weeks of treatment, the primary endpoint, percent change from baseline in hepatic fat fraction measured by MRI-PDFF, will be evaluated. At 36 weeks, all patients will undergo liver biopsy.

Following randomization, patients will begin the initial 12-week Treatment Period. Patients will undergo liver MRI-PDFF prior to the start of treatment (pre-dose), at the end of the initial 12-week Treatment Period (12 weeks after the start of treatment), and after continuing blinded treatment (MGL-3196 or placebo) for an additional 24 weeks (36 weeks after the start of treatment). A liver biopsy will be obtained prior to start of treatment (pre-dose) if the patient has not had a qualifying liver biopsy within 180 days of randomization and at 36 weeks. During the study, patients will return to the study site periodically for assessment of vital signs (temperature, pulse, respiratory rate, and seated blood pressure), 12-lead electrocardiogram (ECG), and clinical laboratory testing (hematology, chemistry, and urinalysis). Blood samples will be collected at specified times for the assessment of lipid parameters, thyroid hormone

parameters, and other biomarkers. Patients will be evaluated for adverse events (AE) and concomitant medication use throughout the study. Patients should be monitored throughout the study for clinical signs and symptoms of hyper- or hypothyroidism.

A PK assessment will be made at Week 2 including a pre-dose fasting determination and assessments at 2, 4, 6, and 8 hours postdose. As described in the PK section, patients who have a calculated AUC_{inf} ≤ to 5500 ng*hr/ml (combined MGL-3196 and MGL-3196-M1) will continue on 80 mg. As described in Section 8, patients (~10%) with an AUC_{inf} ≤3000 ng*hr/ml who are predicted to have exposure at 100 mg that remains AUC ≤5550 based on a conservative calculation, the MGL-3196 dose will be increased from 80 mg to 100 mg at the Week 4 visit. For patients with combined AUC >5500 ng*hr/ml, the dose of MGL-3196 will be reduced to 60 mg per day at the Week 4 visit. Patients with AUC_{inf} >11,000 (combined MGL-3196 and M1) plus an SHBG>150% increase from baseline at Week 2 (indicative that average or steady-state exposure is >11,000 ng*hr/ml) will be downtitrated to 40 mg at Week 4. If SHBG ≤150%, then the patient will be downtitrated to 60 mg at Week 4, and the patient's 4 Week SHBG and predose concentrations will be evaluated. If at Week 4 SHBG >150% increase from baseline and predose >40ng/ml for MGL-3196 and >40ng/ml for M1, then the patient will be down-titrated to 40 mg at the Week 8 visit.

A combined MGL-3196 and MGL-3196-M1 exposure ≤ 5500 ng*hr/ml has not demonstrated a drug-related decrease in FT4 to < LLN < 0.7 ng/dL) in Phase 1 studies. However, patients with NASH without symptoms of hypothyroidism may occasionally have low baseline levels of FT4, with normal FT3 and TSH, and are not clinically or subclinically hypothyroid. Low FT4 in the presence of normal FT3 (active hormone) and TSH is not consistent with hypothyroidism; low FT4 is observed in euthyroid (formerly hypothyroid) patients treated with T3 monotherapy [20]

In patients continuing on 80 mg MGL-3196 (AUC_{inf} < 5500) who have FT4 LLN (>.55 to <.7 ng/dL):

At Week 4 or later:

- If FT3 remains normal (not decreased to <LLN) and TSH remains normal, and the patient does not have symptoms of hypothyroidism, the patient may continue on the current dose.
- Down-titration to 60 mg may be made at the next visit, Week 8 or later if FT4 <LLN (>.55 to <.7 ng/dL), FT3 and TSH are normal, and there is evidence that the decrease in FT4 <LLN (>.55 to <.7 ng/dL) is drug-related including a >30% decrease from baseline in FT4, an increase in SHBG >120% from baseline and predose exposure >40 ng/ml for MGL-3196 and > 40 ng/ml for MGL-3196-M1 that is consistent with increased drug exposure relative to the Week 2 PK assessment.

In patients down-titrated to 60 mg at Week 4 who have FT4 LLN (>.55 to <.7 ng/dL):

- At Week 8 or later, if FT3 and TSH remain normal, the patient does not have symptoms of hypothyroidism, the patient may continue on the current dose.
- Down-titration to 60 or 40 mg may be made at the next visit, Week 12 or later if FT4 <LLN (>.55 to <.7 ng/dL), FT3 and TSH are normal, and there is evidence that the decrease in FT4 <LLN (>.55 to <.7 ng/dL) is drug-related including a >30% decrease

from baseline in FT4, an increase in SHBG >120% from baseline and predose exposure >40 ng/ml for MGL-3196 and > 40 ng/ml for MGL-3196-M1 that is consistent with increased drug exposure relative to the Week 2 PK assessment.

If these conditions are not met, the patient will continue on current dose and thyroid indices reassessed at subsequent visits. Any randomized patients previously down-titrated based on LLN FT4 per Protocol Amendment 2 who did not meet these criteria for down-titration (Amendment 3) will be returned to the previous dose in a blinded fashion at the next visit.

If FT4 is ≤.55 ng/dL on 80 or 60 mg dose with normal FT3 and TSH (which are unblinded to study personnel for all patients), even if the decrease in FT4 is not felt to be primarily drug-related, a down titration to 60 or 40 mg, respectively, may be made at the next visit as determined by Sponsor and Medical Monitor after blinded review. If it is determined after blinded review by Sponsor and Medical Monitor that thyroid therapy should be initiated, because thyroxine cannot be initiated without breaking the blind, the patient will be discontinued from the study and Early Termination assessments will be performed.

A Data Safety Monitoring Board (DSMB) will oversee the study to ensure patient safety and to advise if any dosing alterations are recommended. The DSMB will review safety, including thyroid hormone effects, liver-related events (ie, clinically meaningful elevations in liver enzymes ALT, AST, and bilirubin), cardiovascular assessment (ie, changes in ECG, cardiac biomarkers, cardiac AEs, cardiac symptoms, arrhythmias, etc.), SAEs for trends in causality, and other efficacy (lipid parameters) and safety data as needed. The DSMB will perform regularly scheduled reviews the first will be scheduled when 25 patients have received 6 weeks of treatment. Additional regularly timed meetings will be held and ad hoc meetings may also occur at the discretion of by Madrigal and the DSMB.

The schedule of procedures can be found in Protocol Appendix A.

2.2.1 Extension Study

An open label active treatment 36 Week extension study was conducted in patients completing the Main Study for patients who consented to the Extension study and met the inclusion criterion of significant elevation in liver enzymes at Week 30 of the main study as defined in Protocol Amendment 6. The treatment code was unknown for patients entering the extension study, but treatment code was revealed while the Extension study was ongoing. Patients randomized to MGL-3196 during the Main study remained on the same dose of MGL-3196 or had a prespecified increase in dose upon entering the Extension study. Former placebo patients were treated with 80 mg MGL-3196 initially, then uptitrated to 100 mg, remained on 80 mg or had their dose reduced to 60 mg at Week 4 based on a PK sample at Week 2. During the Extension study, based on additional safety results from the Main study, all Extension study patients had doses adjusted to at least 80 mg per day, and several Extension study patients took 100 or 120 mg per day.

2.3 Study Endpoints

2.3.1 Efficacy Variables

2.3.1.1 MRI-PDFF Assessments

The primary efficacy variable is percent change in hepatic fat fraction by MRI-PDFF from baseline at 12 weeks for MGL-3196 80 mg versus placebo in patients with biopsy-proven NASH.

The other related efficacy variables with MRI-PDFF at Week 12 are:

- Change in hepatic fat fraction by MRI-PDFF from baseline;
- Proportion of patients who have 30% or more reduction in hepatic fat fraction by MRI-PDFF from baseline at Week 12, that is, the percent change from baseline to Week 12 is ≤-30%;
- Proportion of patients who have 5% or more absolute reduction in hepatic fat fraction by MRI-PDFF from baseline at Week 12, that is, the change from baseline to Week 12 is ≤-5% in hepatic fat; and
- Proportion of patients who have 5% or less hepatic fat fraction by MRI-PDFF at Week 12, that is, the hepatic fat fraction measurement at Week 12 is ≤5%.

The efficacy variables at Week 36 include:

- Five efficacy endpoints in hepatic fat fraction by MRI-PDFF as specified in the primary Week 12 efficacy endpoint from baseline at 36 weeks for MGL-3196 80 mg versus placebo;

2.3.1.2 Liver Biopsy Assessments

- NASH Biopsy Responders:
    - The number and percentage of patients achieving a 1-point reduction in NAS;
    - the number and percentage of patients achieving a 2-point reduction in NAS at 36 weeks;
    - the number and percentage of patients achieving a 2-point reduction in NAS at 36 weeks; and either ≥1 point reduction in lobular inflammation or ≥1 point reduction in hepatocellular ballooning; and
    - the number and percentage of patients achieving a 2-point reduction in NAS at 36 weeks; and no worsening of fibrosis.
- NASH Component Responder definition (only requires one or more criteria):
    - reduction in hepatocellular steatosis; or
    - reduction in lobular inflammation; or
    - reduction in hepatocellular ballooning.

- Fibrosis Responders:
  - the number and percentage of patients achieving ≥ 1 stage reduction in fibrosis;
  - the number and percentage of patients achieving ≥ 1 stage reduction in fibrosis; and no worsening of NAS; and
  - the number and percentage of patients achieving ≥ 2 stage reduction in fibrosis.
- NASH Resolution Responder Definition (requires both criteria):
  - lobular inflammation = 0 or 1; and
  - hepatocellular ballooning = 0.
- NASH Resolution Responders with at least a 2-point reduction in NAS at Week 36.
- Other exploratory NASH Resolution Responders:
  - [redacted]
- The number and percentage of patients achieving a change in NASH diagnosis (ie, from definite NASH to not-NASH based on histology) from baseline to Week 36.

Other liver biopsy analyses include:

- [redacted]
- The change in portal inflammation from baseline at 36 weeks; and
- [redacted]
- Correlation of reduction of liver fat on MRI-PDFF at 12 and 36 weeks with improvement or resolution of NASH on liver biopsy.

2.3.1.3 Other safety and efficacy variables include the following:

- The change in thyroid axis hormones from baseline at 12 and 36 weeks for MGL-3196 80 mg versus placebo;
- The percent change from baseline to Week 12 and 36 in SHBG;
- The change in hsCRP, ALT, AST, GGT, direct LDL-C, HDL-C, non-HDL-C, TC, TG, ApoB, Lp(a), ApoCIII, CK-18, FIB-4, Pro-C3, and ELF test from baseline at 12 and 36 weeks for MGL-3196 80 mg versus placebo, including pre-specified baseline cutoffs for ALT, LDL-C, Lp(a) and fibrosis biomarkers;
- The change in HR-QoL SF-36 scores from baseline at 12 and 36 weeks for MGL-3196 80 mg versus placebo;

- The change in glucose, insulin, HOMA-IR, HbA1c, NEFA, adipo-IR, and adiponectin from baseline at 12 and 36 weeks for MGL-3196 80 mg versus placebo;
- The change in body weight, BMI, waist circumference, waist-hip ratio, and blood pressure from baseline at 12 and 36 weeks for MGL-3196 80 mg versus placebo;
- The correlation between plasma MGL-3196 exposure to changes in efficacy and safety biomarkers;
- The correlation between genomic markers of NASH and responses to MGL-3196;
- The correlation in reduction of liver fat on MRI-PDFF at 12 and 36 weeks with improvement or resolution of NASH on liver biopsy; and
- The correlation between percent change in SHBG from baseline to Week 12 and 36 with efficacy and safety measurements at Week 12 and 36 visit.

2.3.2 Pharmacokinetics

Blood samples for PK assessments will be obtained pre-dose from patients at the Baseline Visit (Day 1), Weeks 2, 4, 8, 16, 24, and 36, and the Early Termination Visit. Patients will be instructed not to take study drug prior to PK sampling at these respective study weeks.

At Week 2 visit, PK assessments will be made in the clinic after over-night fast at pre-dose. An 80 mg dose of MGL-3196 or placebo will be administered and PK assessments will be made at 2, 4, 6, and 8 hours post dose. Total AUC_{inf} will be calculated for MGL-3196 and MGL-3196-M1 based on Week 2 $AUC_{(0-8)}$, exposure at 8 h and historic geometric mean elimination rate constant for MGL-3196 and MGL-3196-M1 at the 80 mg dose. Modeling based on historic data obtained for the 80 and 100 mg doses indicate that, using this approach, AUC_{actual} ($AUC_{(0-24)}$) is estimated within 1.8%, SD 5.4% for MGL-3196 given at a dose of 80 mg and is slightly over estimated for MGL-3196-M1 AUC by 5.7% SD 8.5% at a dose of 80 mg. Similar results were obtained for a 100 mg dose. Dose adjustment will be made to 60 mg once per day for any patients whose combined estimated MGL-3196 and MGL-3196-M1 ($AUC_{(0-24)}$) at 2 weeks on the 80 mg dose exceeds 5500 ng*hr/ml. All subjects previously dosed at 50 mg MGL-3196 had combined MGL-3196 plus MGL-3196-M1 less than 5500 ng*hr/ml, geo mean= 3400 ng*hr/ml. Notably, the trough pre-dose drug levels obtained at 50 -100 mg in the MAD study are also predictive of subjects' relative AUCs at a given dose. For patients whose combined estimated $AUC_{(0-24)}$ at 2 weeks at the 80 mg dose is >11,000 ng*hr/ml, plus SHBG>150% increased from baseline, the dose will be decreased to 40 mg at the 4 Week visit (or later per Synopsis), because, even given the greater than dose proportionate increase in exposure between a 60 and 80 mg dose, the exposure at 60 mg may be ≥5500 ng*hr/ml. For patients whose combined estimated $AUC_{(0-24)}$ at 2 weeks at the 80 mg dose is ≤3000 ng*hr/ml, the dose will be increased to 100 mg at the 4 Week visit. As based on conservative estimates, the predicted exposure at 100 mg will be significantly ≤ 5500 ng*hr/ml. This prediction is based on data obtained in healthy volunteers in which there is greater than dose proportionate increases in MGL-3196 between 80 and 100 mg daily doses, and less than dose proportionate increases in MGL-3196-M1 between 80 and 100 mg doses. Pharmacokinetic data (blinded to patient ID) obtained from the first 21 patients randomized to MGL-3196 suggests that there are no clear differences in MGL-3196 or MGL-3196-M1 exposure at the 80-mg dose between healthy volunteers and NASH patients enrolled in this study.

2.3.3 Safety Endpoints

Safety variables to be assessed include safety laboratory tests, vital signs and anthropometrics, 12-lead ECG with rhythm strip, dual-energy x-ray absorptiometry, physical examinations, adverse events, and clinical assessments.

3. STATISTICAL METHODOLOGY

3.1 Baseline, Endpoint, and Other Statistical Considerations

Baseline for all efficacy and safety variables will be defined as Baseline visit (Day 1). If the measurement at this visit is missing, the last measurement prior to the first dose of randomized study drug will be used.

For efficacy variables other than hepatic fat fraction, early termination and unscheduled visits will be assigned a visit number based on the date of assessment relative to the first dose of the double-blind study drug. Visit windows will be defined as the scheduled time from dosing date with a 7-day window (±7 days). Within a visit window, if a scheduled visit occurs, then the measurement from this scheduled visit will be used as the measurement for this visit window. If no scheduled visit occurs within a visit window, the last measurement within the window will be used. If no visits occur within a visit window, the measurement of this visit will be treated as missing.

If the last dose date is missing due to subject lost to follow-up, then it will be assumed the subject is still on study drug at the last clinic visit. For lipids measurements, only data up to one day after the last dose will be considered valid and used in the analyses. For MRI-PDFF and liver biopsy results, the measurements up to three weeks after the last dose will still be considered valid and used in the analyses. For liver enzyme measurements, all the measurements will be used whether the subject is on or off study drug at the clinic visit.

If weight is missing at the Week 12 (or Week 36) visit and the subject completed the Week 12 (or Week 36) visit, the average of the weight at the previous scheduled visit and next scheduled visit will be used as the Week 12 (or Week 36) measurement.

Descriptive statistics (number of patients, mean, standard deviation, median, minimum, and maximum) will be used to summarize the continuous data. Discrete measures will be summarized using counts and percentages.

3.2 Analysis Populations

3.2.1 Intent-to-Treat Population

The Intent-to-Treat (ITT) Population will include all patients who are randomized in the study and receive at least 1 dose of study drug.

3.2.2 Modified Intent-to Treat Population

The modified intent-to-treat (mITT) population includes all patients who were randomized in the study, received at least 1 dose of study drug, and had lipid and other efficacy measurements at Week 4 or later visits. The mITT population will be used for all secondary efficacy analyses.

3.2.3 MRI-PDFF Evaluable Population

The MRI-PDFF Evaluable population includes all patients who were randomized in the study, received study drug, finished Week 12 Visit, with valid MRI-PDFF measurements at both baseline and Week 12 Visit.

3.2.4 Liver Biopsy Evaluable Population

The liver biopsy evaluable population includes all patients who were randomized in the study, received study drug, with two valid evaluable liver biopsies, one at baseline and one post-treatment. In addition to the following per-protocol populations, additional analyses on the liver biopsy evaluable population will be performed using a Posthoc Per-Protocol Population, as warranted.

3.2.5 12-Week Per Protocol Population

The 12-Week Per Protocol Population will include all MRI-PDFF Evaluable patients who do not have any major protocol deviations during the first 12 weeks. The following criteria will be evaluated for major deviations prior to unblinding of the treatment allocation:

- No eligibility criterion violations;
- Did not discontinue study drug prior to Week 12;
- Patients who are <=75% compliant over the course of the initial 12-week Treatment Period and/or the period between Week 8 and Week 12;
- Not taken any prohibited concomitant medication during 12-week Treatment Period;
- Not unblinded during the 12-week double-blind treatment period; and
- No other substantial protocol violations during 12-week Treatment Period.

The 12-Week Per Protocol Population will be used to assess robustness of the analysis of hepatic fat fraction results of MRI-PDFF during the 12-week Treatment Period.

3.2.6 36-Week Per Protocol Population

The 36-Week Per Protocol Population will include all MRI-PDFF Evaluable patients who have a Week 36 MRI-PDFF measurement, and a baseline and Week 36 liver biopsy.

The 36-Week Per Protocol Population will be used to assess robustness of the analysis of hepatic fat fraction results of MRI-PDFF during the 36-Week period.

Additional potential patients may be excluded from Week 36 Per Protocol Population if the patients:

- Take any prohibited concomitant medication during 12-36 week Treatment Period; or
- Patients who are <=75% compliant over the course of the 12-36 week Treatment Period and/or the period between Week 30 – 36.

Please note that non-compliance from Week 30-36 is not per protocol for Week 36 laboratory tests and Week 36 MRI-PDFF; non-compliance between Week 30-36 with drug compliance for Weeks 12-36 >75% are per protocol for liver biopsy.

For patients who are in the Week 36 Per Protocol Population with Week 36 liver biopsy measured more than two weeks of last dose date, these liver biopsy values will not be used in the Week 36 Per Protocol analysis for liver biopsy responders.

3.2.7 Safety Population

The Safety Population will include all patients who are randomized in the study and receive at least 1 dose of study drug and will be used for all safety analyses. The Safety Population is identical to the ITT Population.

3.3 Patient Disposition

The reason given for each screen failure will be summarized and listed. Patient disposition will be provided for all randomized patients. The number and percentage of patients in each of the following disposition categories will be presented:

- Patients who are randomized,
- Patients who start study drug,
- Patients who have PK assessment at Week 2,
- Patients who complete the 12-week Treatment Period,
- Patients who complete the 36-week study,
- Patients who withdraw from the study during the 12-week Treatment Period, and
- Patients who withdraw from the study.

For randomized patients who withdraw from the study, the primary reason for the withdrawal will be listed and summarized.

The number of patients included in each analysis population will be summarized.

3.4 Demographics and Baseline Characteristics

Demographics (age, sex, race, and ethnicity), body weight, height, and BMI will be summarized by randomized treatment using descriptive statistics for the randomized patients based on the Safety Population. Age group (<65 years versus ≥65 years), age group by median (≥median versus <median), and BMI category (≥30 kg/m^2 versus <30 kg/m^2) will also be summarized. Other disease characteristics include diabetes and hypertension status.

Baseline levels of FT4, FT3 and TSH will be summarized to evaluate hypothyroidism. Baseline liver enzymes (ALT, AST), and NAS from liver biopsy will also be summarized for disease characteristics, as well as PNPLA3 and baseline fibrosis score category (0/1 and 2/3).

Medical history will be coded using the Medication Dictionary for Regulatory Activities (MedDRA). The number and percentage of patients with medical history will be summarized by their MedDRA preferred term within system organ class and by treatment.

3.5 Prior/Concomitant Medications

The use of any prior medication and concomitant medication will be listed for the Safety Population with an indication of whether the medication was prior or concomitant during the study.

Medications will be coded using the most recent version of World Health Organization (WHO) Drug Dictionary. The number and percentage of patients taking each concomitant medication will be summarized by preferred term within Anatomical Therapeutic Chemical (ATC) classification and by treatment group for the Safety Population. Concomitant medication is defined as any medication taken on or after the first dose day of study medication.

3.6 Study Medication Exposure and Compliance

Days of exposure to study medication will be summarized by treatment group for the Safety Population using descriptive statistics. Exposure in days is defined as the date of last dose of study drug minus the date of first dose of study drug plus 1. In addition, a contingency table will be provided to display the number and percentage of patients with exposure in the following categories:

- 1 to ≤2 weeks (1-14 days),
- >2 to ≤4 weeks (15-28 days),
- >4 to ≤8 weeks (29-56 days);
- >8 to ≤12 weeks (57-84 days),
- >12 to ≤36 weeks (84-252 days), and
- >36 weeks (253 days or more).

Compliance will be calculated based on the total actual number of capsules taken as compared to the total expected number of capsules taken (2 per day). Percent compliance with the study medication during the 12-week Treatment Period (from Week 4 through Week 12) and the whole 36-week Treatment Period (from Week 4 through Week 36) will be summarized by treatment group for the Safety Population using descriptive statistics. Additionally, the number and percentage of patients within each treatment group with overall compliance in the following categories: <80%, 80% to 120%, and >120%, will be provided.

All the patients randomized to the active study drug group will take MGL-3196 80 mg. Based on the PK assessment and SHBG at Week 2, the dose at Week 4 may stay at 80 mg, it may be increased to 100 mg, or it may even be reduced to 60 mg. At Week 8 visit, the dose may be reduced to 40 mg. Duration of each dose level will be summarized, and number/percentage of patients taking each dose at each visit will be tabulated.

3.7 Analysis of Efficacy

3.7.1 Descriptive Statistics

All available raw and derived data will be listed for the randomized patients. Descriptive statistics of the baseline, change from baseline and/or percent change from baseline to each post-baseline visit will be presented by treatment group for each efficacy measurement for the ITT population.

3.7.2 Analyses of Hepatic Fat Fraction by MRI-PDFF

Summary statistics (number of patients, mean, standard deviation, median, minimum, and maximum) for hepatic fat fraction by MRI-PDFF at all visits and percent change from baseline will be provided for the MRI-PDFF Evaluable Population.

Analyses of hepatic fat fraction will be performed based on the MRI-PDFF Evaluable population and repeated based on the 12-Week Per Protocol population.

The primary efficacy endpoints are summarized and analysed by treatment group (MGL-3196 vs. placebo) by overall population and subgroups. For overall analysis and certain subgroups, the primary efficacy endpoints will also be analysed by MGL-3196 exposure level. The

MGL-3196 exposure levels are determined by the plasma MGL-3196 concentration at Week 2 and later visits to represent the overall MGL-3196 exposure level for patients.

3.7.2.1 Analyses of Percent Change

Percent change in hepatic fat fraction = 100 X (Week 12 – Baseline) / Baseline.

An analysis will be performed on the percent change from baseline to Week 12 in hepatic fat fraction by MRI-PDFF. An analysis of covariance (ANCOVA) model with treatment as a factor will be analyzed. The SAS sample code is listed:

```
*********************************************************************
** PCPDFF = Percent change in MRI-PDFF from Baseline to Week 12  **
** Trt = Treatment group                                           **
*********************************************************************
PROC GLM data=EFF;
   CLASS Trt;
   MODEL PCPDFF = Trt;
   LSMEANS Trt / e diff cl;
RUN;
```

In this analysis, the treatment group will be either MGL-3196 active or placebo. A pairwise comparison will be performed between MGL-3196 active and placebo. LS means for the treatment difference, 95% confidence interval and p-value will be provided.

3.7.2.2 Analyses of Absolute Change

Absolute change in hepatic fat fraction = Week 12 – Baseline.

An analysis will be performed on the absolute change from baseline to Week 12 in hepatic fat fraction by MRI-PDFF. An analysis of covariance (ANCOVA) model with treatment as a factor and baseline value as a covariate will be analyzed. The SAS sample code is listed:

```
*********************************************************************
** CPDFF = Change in MRI-PDFF from Baseline to Week 12           **
** Trt = Treatment group                                         **
** Base = Baseline value                                         **
*********************************************************************
PROC GLM data=EFF;
   CLASS Trt;
   MODEL CPDFF = Trt Base;
   LSMEANS Trt / e diff cl;
RUN;
```

In this analysis, the treatment group will be either MGL-3196 active or placebo. A pairwise comparison will be performed between MGL-3196 active and placebo. LS means for the treatment difference, 95% confidence interval and p-value will be provided.

3.7.2.3 Week 12 Hepatic Fat Fraction Treatment Goals

There are three treatment goal endpoints that will be used to evaluate the treatment effect.

- % Change from baseline to Week 12 < -30% (% reduction of 30% or more),

- Change from baseline to Week 12 < -5% (absolute reduction of 5% or more), and
- Week 12 hepatic fat fraction <5%.

For each endpoint, a logistic regression model with treatment as a factor and baseline as a covariate will be used to compare the percentages of patients meeting the following treatment goals between the MGL-3196 versus placebo. The odds ratio, confidence interval, and p-value from the logistic regression will be provided.

3.7.2.4 Subgroup Analyses for Week 12 Primary Efficacy Endpoints

The analyses of percent change and absolute change in hepatic fat fraction will be performed for the MRI-PDFF Evaluable population for the following subgroups:

- Gender (male, female);
- Weight loss from Screening (<5% weight loss at Week 12, ≥5% weight loss at Week 12);
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino);
- Type 2 diabetes (yes or no);
- PNPLA3 SNP phenotype (homozygous, heterozygous, negative);
- Fibrosis Stage [0-1 (Total), 2-3];
- NAS Score (≤4, >4);
- Week 12 SHBG group (≥75%, <75% change from baseline to Week 12);
- Baseline liver fat percentage (<20%, ≥20%); and
- Exposure group (AUC <2700 ng*hr/mL, AUC ≥2700 ng*hr/mL).

3.7.2.5 Analyses for Week 36 Primary Efficacy Endpoints (MRI-PDFF)

Analyses of hepatic fat fraction from Baseline to Week 36 will follow the same analysis as Week 12, based on the mITT population. This analysis will also be repeated for the 36-Week Per Protocol Population. If there are patients who early withdraw between Weeks 12 and 36, then the following two analyses will be conducted:

- Analyze the primary endpoint (percent or absolute changes in hepatic fat fraction by MRI-PDFF) without the early withdrawals, that is, only include patients who finish Week 36 MRI-PDFF (mITT Population – Observed); and
- Last observation carried forward between Weeks 12 and 36, and use this last observation as the Week 36 value for the analysis (mITT Population – LOCF).

The percent and absolute change in hepatic fat fraction by MRI-PDFF between Week 12 and Week 36 without the early withdrawals will be analyzed, that is, only include patients who finish Week 36 MRI-PDFF. The results between Weeks 12 and 36 will show durability of response and whether MRI-PDFF at week 12 is a good a predictor of response or Week 36 is better.

Additional sensitivity analysis may be performed with other statistical models, such as Mixed effect Model Repeat Measurement (MMRM) method.

3.7.2.6 Subgroup Analyses for Week 36 Primary Efficacy Endpoints

The analyses of percent change and absolute change in hepatic fat fraction will be performed for the MRI-PDFF Evaluable population for the following subgroups:

- Week 36 SHBG group (≥88%, <88% change from baseline to mean SHBG post Week 12; and,
- Exposure group (AUC <2700 ng*hr/mL, AUC ≥2700 ng*hr/mL).

3.7.3 Analyses of Liver Biopsy Results

All endpoints will be summarized and analyzed based on the mITT population for MGL-3196 80 mg versus placebo. For liver biopsy, mITT includes only patients with both eligible baseline and Week 36 biopsies, Since this is for exploratory purpose, no multiplicity will be controlled.

Responder variables at Week 36 include:

- NASH Biopsy Responders:
  - The number and percentage of patients achieving a 1-point reduction in NAS;
  - the number and percentage of patients achieving a 2-point reduction in NAS at 36 weeks;
  - the number and percentage of patients achieving a 2-point reduction in NAS at 36 weeks; and either ≥1 point reduction in lobular inflammation or ≥1 point reduction in hepatocellular ballooning; and
  - the number and percentage of patients achieving a 2-point reduction in NAS at 36 weeks; and no worsening of fibrosis.
- NASH Component Responder definition (only requires one or more criteria):
  - reduction in hepatocellular steatosis; or
  - reduction in lobular inflammation; or
  - reduction in hepatocellular ballooning.
- Fibrosis Responders:
  - the number and percentage of patients achieving ≥ 1 stage reduction in fibrosis;
  - the number and percentage of patients achieving ≥ 1 stage reduction in fibrosis; and no worsening of NAS; and
  - the number and percentage of patients achieving ≥ 2 stage reduction in fibrosis.
- NASH Resolution Responder Definition (requires both criteria):
  - lobular inflammation = 0 or 1; and
  - hepatocellular ballooning = 0.

- NASH Resolution Responders with at least a 2-point reduction in NAS at Week 36.
- Other exploratory NASH Resolution Responders:
  - [redacted]
- The number and percentage of patients achieving a change in NASH diagnosis (ie, from definite NASH to not-NASH based on histology) from baseline to Week 36.

The definite NASH based on histology is specified as: fibrosis stage 1 to 3 and a NAS of ≥4 with at least a score of 1 in each of the following NAS components:

- Steatosis (scored 0 to 3),
- Ballooning degeneration (scored 0 to 2), and
- Lobular inflammation (scored 0 to 3).

Fisher's exact test will be used to compare the percentages between MGL-3196 80 mg versus placebo. Confidence intervals and p-values will also be provided.

NOTE:

- The analyses performed on the NASH Resolutions endpoint will exclude patients with >9.5% weight loss from Screening to Week 36 (the percent change in body weight from Screening to Week 36 is >-9.5%), including the patients only as secondary analyses. Conversely, other liver biopsy and study endpoints will include the patients, excluding them only as secondary analyses.

Other liver biopsy variables include:

- The change in fibrosis stage from baseline at 36 weeks; as a continuous variable and descriptively by numbers and percentages of patients in which fibrosis stage improved, was unchanged or worsened, for exploratory purposes only
- The change in portal inflammation from baseline at 36 weeks; and
- The change in parameters of NASH liver biopsy (NAS) from baseline at 36 weeks. The NAS score will also be summarized descriptively as a continuous variable for exploratory purposes only.
- Correlation of reduction of liver fat on MRI-PDFF at 12 and 36 weeks with improvement or resolution of NASH on liver biopsy.

Analyses of liver biopsy data will be performed on the mITT population for the following subgroups:

- Week 36 SHBG group (≥88%, <88% change from baseline to mean SHBG post Week 12 through Week 36);
- Exposure group (AUC <2700 ng*hr/mL, AUC ≥2700 ng*hr/mL);
- Baseline Fibrosis Stage 2/3;

- Weight loss from Screening to Week 36 (<5% weight loss, ≥5% weight loss, <7% weight loss, ≥7% weight loss, >9.5% weight loss);
- Baseline MRI-PDFF ≥20%;
- Gender (male, female);
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino);
- Stage of NASH (≥5, <5);
- PNPLA3 SNP phenotype (Homozygous, Heterozygous, negative);
- Type 2 diabetes (yes or no);
- MGL-3196 treated patients who have 30% or more reduction in hepatic fat fraction by MRI-PDFF from baseline at Week 12 and 36, that is, the percent change from baseline to Week 12 and 36 is ≤-30% versus all Placebo treated patients;
- MGL-3196 treated patients who have 5% or more absolute reduction in hepatic fat fraction by MRI-PDFF from baseline at Week 36, that is, the change from baseline to Week 12 is ≤-5% in hepatic fat versus all Placebo treated patients;
- MGL-3196 treated patients who have 5% or less hepatic fat fraction by MRI-PDFF at Week 12 and 36, that is, the hepatic fat fraction measurement at Week 12 and 36 is ≤5% versus all Placebo treated patients; and
- Any other data driven groups.

Other potential subgroups may be explored if any imbalance exists between treatment groups. Imbalance will be assessed using a chi-square test for categorical data, and an analysis of variance (ANOVA) will be used to test the imbalance for quantitative data.

3.7.4 Analyses of Other Efficacy Endpoints

Summary statistics (number of patients, mean, standard deviation, median, minimum, and maximum) for all visits, including the change from baseline and percent change from baseline, will be provided for each parameter for the mITT population.

The parameters will be analysed based on the mITT population. For certain efficacy endpoints, such as liver enzymes at Week 12, analysis will be repeated on MRI population as well. Since this is for exploratory purpose, no multiplicity will be controlled. Where stated, the last observation carried forward will be used for missing value imputation when applying an ANCOVA model.

3.7.4.1 Liver Enzyme (ALT, AST, and GGT)

The same ANCOVA model for the percent change and change in hepatic fat fraction will be used for analyzing the percent change and change from baseline to Week 12 in ALT, AST, and GGT. That is, the percent change and change from baseline to Week 12 in ALT or AST will be analyzed with the ANCOVA model with treatment as a factor. Baseline ALT, AST, or GGT value will be a covariate in the change from baseline analysis. A pairwise comparison will be

performed between MGL-3196 active and placebo. LS means for the treatment difference, 95% confidence interval and p-value will be provided.

These analyses will be repeated at the Week 30, Week 36 and Week 38 time points for the mITT population. From the results of Week 36 and 38, the sustainability of drug effect will be evaluated.

Moreover, the analyses will also be repeated for the following derived values:

- The lowest ALT, AST, or GGT endpoint value from Week 30 and Week 36. If a patient terminated early, the lowest value from the early termination visit and the previously scheduled visit will be considered the lowest value; and
- The lowest ALT, AST, or GGT value after Week 12 and through Week 36.

Unscheduled visits will not be included in these derivations.

Analyses of ALT will be repeated for females with baseline ALT ≥30 U/L and males with baseline ALT ≥45 U/L.

The Mixed Model for Repeated Measurements (MMRM) will be used to test the robustness of the results of treatment comparison at each visit. The treatment difference will be estimated from the model for Weeks 12, 30 and 36.

Analyses of ALT, AST, and GTT may be performed for the following subgroups:

- Week 12 SHBG group (≥75%, <75% change from baseline to Week 12) for men and women separately;
- Week 36 SHBG group (≥88%, <88% change from baseline to mean SHBG post Week 12 through Week 36);
- Exposure group (AUC <2700 ng*hr/mL, AUC ≥2700 ng*hr/mL);
- Females with baseline ALT ≥30 U/L and males with baseline ALT ≥45 U/L; and
- Any other data driven groups.

3.7.4.2 Lipid Measurements

All lipid measurements will be summarized by visit, as well as absolute and percent change from baseline to each post-baseline visit.

The same ANCOVA model as primary efficacy analysis will be used for analyzing the percent change and change:

- From baseline to Week 4,
- From baseline to Week 12 with LOCF (after Day 1 and up to Week 12),
- From baseline to Week 30 with LOCF (after Week 12 and up to Week 30),
- From baseline to Week 36 with LOCF (after Week 12 and up to Week 36),
- Average between Week 30 and Week 36, and
- Average between Week 16 and Week 36

for direct LDL-C, triglycerides, Lp(a), HDL, Non-HDL, Total Cholesterol, Apo C-III (after Week 12 only), and Apo B.

Lp(a) will be analyzed separately for baseline value ≤10 nmol/L and baseline value >10 nmol/L. In the ANCOVA model for change from baseline, baseline value will be included as a covariate. But the covariate of baseline value will not be included in the model for the analysis of percent change from baseline.

If skewness from normality is observed for triglycerides data, then the triglycerides value will be logarithm-transformed before fitting the ANCOVA model. The treatment difference between two visits will then be transformed back as the percent change from. If the normality assumption still can't be satisfied with the logarithm-transformed data, a non-parametric approach (Wilcoxon rank-sum test) will be used to compare the difference in median TG change or percent change between treatment groups.

A logistic regression model with treatment as a factor and baseline as a covariate may be used to compare the percentages of patients who achieve a change from baseline to endpoint in Lp(a) (≤10 nmol/L vs. >10 nmol/L) between the MGL-3196 dose versus placebo. The odds ratio, confidence interval, and p-value will be provided.

Analyses of lipids may be performed for the following subgroups:

- Week 12 SHBG group (≥75%, <75% change from baseline to Week 12) for men and women separately;
- Week 36 SHBG group (≥88%, <88% change from baseline to mean SHBG post Week 12 through Week 36);
- Exposure group (AUC <2700 ng*hr/mL, AUC ≥2700 ng*hr/mL);
- Baseline Direct LDL Category (<100 mg/dL, ≥100 mg/dL);
- Baseline Triglycerides Category (≤150 mg/dL, >150 mg/dL); and
- Any other data driven groups.

3.7.4.3 Thyroid Axis Hormones

Summary statistics (number of patients, mean, standard deviation, median, minimum, and maximum) for all visits, including the change and % change from baseline, will be provided for TSH, T4, FT4, T3, FT3, TBG, and Reverse T3 for the mITT population.

The same ANCOVA model will be used for analyzing the percent change and change from baseline to Week 12 for Reverse T3 for the mITT population.

- Sex hormones SHBG, testosterone, free testosterone, LH and FSH will be reported by gender and statistical assessments (CFB, %CFB) at Weeks 12 and 36 provided

3.7.4.4 Fibrosis Markers

Fibrosis markers will be assessed by visits based on ITT population. Change and percent change from baseline to Week 12 and Week 36 endpoints will be summarized. The same ANCOVA model will be used for analyzing the change and percent change from baseline to

Week 12 and to Week 36 for fibrosis markers, by overall treatment group, exposure group, and SHBG groups.

Summary and analysis of fibrosis markers will be repeated on all ITT population with Week 12 liver fat reduction ≥30% from baseline in MRI-PDFF for active treatment group vs. all placebo subjects. In addition, the following subgroups will be explored at both Week 12 and Week 36 for each corresponding fibrosis marker:

- Week 12 SHBG group (≥75%, <75% change from baseline to Week 12) for men and women separately;
- Week 36 SHBG group (≥88%, <88% change from baseline to mean SHBG post Week 12 through Week 36);
- Exposure group (AUC <2700 ng*hr/mL, AUC ≥2700 ng*hr/mL);
- Baseline ELF ≥ 9;
- Baseline PIIINP ≥ 8 μg/L;
- Baseline TIMP ≥ 129 μg/L;
- Baseline HA ≥ 50 μg/L; and
- Baseline PRO-C3 ≥ 10 μg/L, and ≥ 17.5 μg/L.

For the fibrosis markers (e.g. ELF, pro-C3, adiponectin and CK-18), change from Week 12 to Week 36 will also be summarized and analyzed. This will show durability of response and whether the biomarkers at week 12 is a good a predictor of response or Week 36 is better, and also provide the significance between drug treated and placebo in the change in non-invasive biomarker parameters between Week 12 and Week 36. If there is a significant decrease in ELF and components in the whole population between Weeks 12 and 36 that is larger than the decrease between baseline and Week 12, it might suggest that differences between Week 12 to Week 36 and Week 0 to Week 36 of fibrosis marker data may be used to predict response on liver biopsy.

3.7.4.5 Other Efficacy Variables

Other efficacy variables will also be summarized as appropriate. These efficacy variables include: Direct and Total Bilirubin, ALP, fibrinogen, hsCRP, weight, BMI, eGFR, Troponin I, HR-QoL SF-36 scores, HOMA-IR, HbA1c, NEFA, adipo-IR, adiponectin, Haptoglobin, waist circumference, waist-hip ratio, and blood pressure.

Change or percent change from baseline to Week 12 and Week 36 endpoints will be summarized. For certain endpoints, Mixed Model for Repeated Measurements (MMRM) will also be used to test the robustness of the results. The MMRM analysis does not impute any missing values. For all endpoints, the active MGL-3196 doses will be grouped into one to be compared with placebo.

In the evaluations of insulin resistance parameters such as glucose, HgbA1C, insulin, HOMA-IR, any other calculated insulin resistance indices, any diabetic patients or patients on diabetes medicines will be excluded, since these parameters are known to be confounded with anti-diabetic treatment. However, this rule will not be applicable to the assessment of adiponectin.

3.7.5 Analyses of Correlation between MRI-PDFF and Other Efficacy Endpoints

The correlation analyses will be conducted for MGL-3196 and placebo group separately. Both percent change and absolute change from baseline to Week 12 will be used for the correlation analyses.

Spearman's rank correlation coefficient (Spearman's ρ) will be used to evaluate the following relationship:

- The correlation between percent change or absolute change in MRI-PDFF and weight loss;
- The correlation between percent change or absolute change in MRI-PDFF and lipid change (percent change or absolute change in direct LDL or triglycerides);
- The correlation between percent change or absolute change in MRI-PDFF and change in ALT;
- The correlation between percent change or absolute change in MRI-PDFF and percent change in SHBG from baseline to Week 12; and
- The correlation in reduction of liver fat on MRI-PDFF at 12 and 36 weeks with improvement or resolution of NASH on liver biopsy.

3.7.6 Analyses of Correlation between MGL-3196 Exposure and Efficacy Endpoints

Analyses will also be conducted to investigate the correlation between MGL-3196 exposure with efficacy and safety endpoints.

3.8 Analysis of Safety

The safety endpoints for this study include: AEs, safety laboratory assessments, vital signs, 12-lead ECGs, bone mineral density by Dexa scan, and clinical assessments. The safety endpoints will be summarized based on the Safety Population.

3.8.1 Adverse Events

A Treatment-Emergent Adverse Event (TEAE) is defined as any AE that occurred for the first time after the first dose of double-blind study drug or existed prior to the first dose and worsened during the post dosing period. AE worsening applies to severity, or relationship to study drug.

A summary overview of TEAEs will be provided, which presents the number and percentage of patients in each treatment group from the Safety Population for the 12-week Treatment Period and the whole 36-week Treatment Period satisfying each of the following categories:

- Any TEAEs,
- Maximum severity of TEAEs,
- Study drug-related TEAEs,
- Maximum severity of study drug-related TEAEs,
- All TE-SAEs,
- All study drug-related TE-SAEs,
- TEAEs leading to death,
- TEAEs leading to study drug discontinuation, and
- Study drug-related TEAEs leading to interruption/study drug discontinuation.

Adverse events will be coded using MedDRA. The number and percentage of patients with TEAEs will be summarized by their MedDRA preferred term within system organ class and by treatment. AEs will be counted by the number of events as well as the number of patients. For event count summaries, multiple AE events with the same MedDRA coded terms (preferred term and system organ class) and onset date and time from the same patient will only be counted once. For patient count summaries, multiple AE events with the same MedDRA coded terms (preferred term and system organ class) from the same patient will only be counted once. Whether the AE belongs to the 12-week Treatment Period and the whole 36-week Treatment Period will be determined by the AE onset date.

The number and percentage of patients with TEAEs will be summarized by reported maximum severity within each MedDRA preferred term within system organ class and by treatment.

The number and percentage of patients with any TE-SAEs, drug-related TEAEs, drug-related TE-SAEs, and TEAEs and TE-SAEs that lead to study drug discontinuation will be summarized by preferred term within system organ class and by treatment. In these summaries, any patients reporting multiple episodes of the same TEAE (i.e., same preferred term and system organ class) will be counted once.

All SAEs will be listed with an indication of whether the SAE was treatment emergent or started prior to treatment.

All TEAEs that are reported as having led to a withdrawal of study drug will be listed.

3.8.2 Clinical Laboratory Assessments

3.8.2.1 Chemistry, Hematology, and Other lab assessments

Descriptive statistics of each chemistry parameter, hematology parameter, thyroid axis hormone, coagulation parameter, sex hormone, metabolic panel parameter, and other markers will be presented for baseline values and for values and changes from baseline at each post-baseline visit. These will be presented by treatment group for each parameter using the Safety Population. SHBG will be summarized overall and also by gender.

Counts and percentages of patients with any post-baseline observation that is below the lower limit of normal (<LLN) or above the upper limit of normal (>ULN) will be summarized for each chemistry and hematology parameter by treatment group and overall. Standard laboratory cutoffs will be used.

3.8.3 12-Lead Electrocardiogram

ECG parameters (heart rate, PR interval, QRS interval, QT interval, QTcB interval, QTcF interval, RR Interval, and overall interpretation) will be summarized by randomized treatment using descriptive statistics for the Safety Population. The average of 3 readings will be used in the summary. For the overall interpretation, the most severe interpretation at the visit will be summarized.

3.8.4 Physical Examination

The number and percentage of patients with abnormal physical examination findings will be presented for the Safety Population.

3.8.5 Vital Signs, Weight, Height, BMI

Vital signs parameters will be summarized using descriptive statistics for the Safety Population. The change from baseline will also be presented.

3.8.6 Bone Mineral Density

Bone mineral density by DEXA scan will be summarized at baseline and Week 36 for the Safety Population. The change from baseline will also be presented.

3.8.7 Other Safety Parameters

Other safety data will be listed.

3.9 Pharmacokinetic Analysis

The PK data is collected at:

- Week 2 PK data collected at pre-dose, and at 2, 4, 6, and 8 hours postdose for MGL-3196 and MGL-3196-M1;
- Pre-dose MGL-3196 and MGL-3196-M1 at weeks 4, 8, 16, and 24.

The PK parameter $AUC_{(0\text{-}8hr)}$ will be calculated for MGL-3196 and MGL-3196-M1, respectively, with non-compartment method. Linear up and linear down calculation will be used for the areas of trapezoids. $AUC_{(0\text{-}8hr)}$ is calculated using the exposure data from the study time points, predose (0 hour), and 2, 4, 6, 8 hours postdose.

If pre-dose concentration of MGL-3196 or MGL-3196-M1 is not measurable (BLQ), then the LLOQ value from bioanalytical lab will be used. (1 ng/mL for both MGL-3196 and MGL-3196-M1).

The overall exposure, AUC_{inf}, will be calculated as $AUC_{(0\text{-}8hr)}$ and the area of the terminal phase after 8 hours. It's calculated as:

$$AUC_{inf} = AUC_{(0\text{-}8hr)} + C_{8hr} / \lambda.$$

The elimination rate constant (λ) that will be used is the same for every patient. For MGL-3196 (80 mg) the elimination rate constant is $0.292\ hr^{-1}$. and for MGL-3623 it is $0.191\ hr^{-1}$. These are the geometric means elimination rate constant from the MAD study which were provided from Madrigal.

The combined AUC is calculated as: AUC_{inf} of MGL-3196 + AUC_{inf} of MGL-3196-M1.

PK concentrations and PK AUC parameters will be summarized descriptively for patients who are taking MGL-3196 active doses.

3.10 Interim Analysis

Interim analysis is not designed for this study. However, an analysis will be performed once all patients reach Week 12 of the Treatment Period. The analysis will assess the efficacy of MGL 3196 on percent change in liver fat by MRI-PDFF.

4. SAMPLE SIZE DETERMINATION

Approximately 117 patients in total will be randomized to 80 mg (2/3) or placebo (1/3) treatments. After PK assessment, a significant fraction of patients' combined AUC (MGL-3196 plus MGL-3196-M1) may exceed 5500 ng*hr/ml resulting in a dose reduction to 60 mg. In order to power the study appropriately for an efficacious dose (which may be ≥80mg), randomization may continue until approximately 36 patients have continued at Week 4 on the 80 mg dose. It is estimated that the treatment difference of percent change in hepatic fat fraction from baseline to Week 12 between any dose of MGL-3196 and the placebo group is about -30%. With a common standard deviation for the percent change in hepatic fat fraction at 35%, 36 patients per group completing the Week 12 visit will provide 90% power with a 2-sample t-test. The significance level is set as 0.025 for the consideration of multiplicity in comparisons of 2 daily dosing regimens of MGL-3196 versus the placebo group. The enrollment size is designed to allow for 10% drop out before the Week 12 visit, and as such, patients who drop out of the study will not be replaced.

With the sample size of 39 patients per treatment, the power for selected secondary efficacy endpoints was calculated. The baseline hepatic fat fraction is expected to be approximately 15% to 17%, thus the change from baseline to Week 12 in hepatic fat fraction is approximately -5.5% when comparing to placebo. With common standard deviation of change in hepatic fat fraction as 6%, 36 completed patients will provide 94% power for the study when evaluating the treatment effect in change of hepatic fat fraction.

In addition, the sample size is expected to provide meaningful liver biopsy-related data.

5. PROGRAMMING SPECIFICATIONS

Statistical analyses will be performed using SAS® (Cary, NC) version 9.3 or above. All available data will be presented in patient data listings, which will be sorted by site number, unique patient identifier and where appropriate, visit number and visit/assessment date.

The programming specifications, including mock-up validity listings, analysis tables, figures, and data listings, as well as the derived database specification, will be prepared in stand-alone documents. The programming specifications document will be finalized prior to the database lock and unblinding the treatment code.

6. REFERENCES

Jonklaas J, Bianco AC, Bauer AJ, Burman KD, Cappola AR, Celi FS, Cooper DS, Kim BW, Peeters RP, Rosenthal MS, Sawka AM; *American Thyroid Association Task Force on Thyroid Hormone Replacement. Guidelines for the treatment of hypothyroidism: prepared by the american thyroid association task force on thyroid hormone replacement.* Thyroid. 2014 Dec;24(12):1670-751. doi: 10.1089/thy.2014.0028

Glynn R., Laird N., Rubin DB (1986). *Selection Modeling Versus Mixture Modeling with Nonignorable Nonresponse. Drawing Inferences from Self-Selected Samples*. pp 115-142.

Hogan J. W., Roy J., and Korkontzelou C.. *Handling Drop-out in Longitudinal Studies.* Statistical Medicine, 23:1455–1497. 2004;

Little R.. (1993) *Pattern-Mixture Models for Multivariate Incomplete Data. Journal of the American Statistical Association*. Vol. 88, No. 421 (Mar., 1993), pp. 125-134

Little, R., Yau, L. *Intent-to-Treat Analysis for Longitudinal Studies with Drop-Outs.* Biometrics, 1996, vol. 52, 1324-1333.

Ratitch B.; and O'Kelly M.; *Implementation of Pattern-Mixture Models Using Standard SAS/STAT Procedures*. PharmaSUG2011 - Paper SP04, 2011

Ratitch B, O'Kelly M and Tosiello R. Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. *Pharmaceutical Statistics* 2013; **12** 337-347

Rubin, D.B. *Multiple Imputation for Nonresponse in Surveys,* New York: John Wiley & Sons, Inc. 1987

Tangen, C.M. and Koch, G.G.. *Nonparametric Analysis of Covariance for Hypothesis Testing with Logrank and Wilcoxon scores and Survival-rate estimation in a Randomized Clinical Trial.* Journal of Biopharmaceutical. Statistics, 9:2, 307-338, 1999.

Yang, Y.. *Sensitivity Analysis in Multiple Imputation for Missing Data.* SAS Global Forum 2014-SAS270, 2014.