Cirius Therapeutics, Inc.

Protocol No.: MSDC-0602K-C009NASH

A Phase 2, Randomized, Double-Blind, Placebo-Controlled, 12-Month, Multiple-Dose Study to Evaluate the Safety, Tolerability and Efficacy of Three Dose Levels of MSDC-0602K in Patients with NASH (EMMINENCE™)

Study ID: MSDC-0602K-C009NASH

STATISTICAL ANALYSIS PLAN

Version: Final 1.5
Date of Issue: 16-May-2019

Author: Stefanie Senger, PhD

Cirius Therapeutics, Inc.
12651 High Bluff Drive
Suite 150
San Diego, CA 92130

161 East Michigan Avenue, 4th Floor
Kalamazoo, MI 49007

Momentum Research Inc.
3100 Tower Blvd., Suite 802
Durham, NC 27707

APPROVALS

The undersigned agree that all required reviews of this document are complete and approve this Statistical Analysis Plan as final.

Signature	Date
Howard C. Dittrich, MD Chief Medical Officer, Cirius Therapeutics, Inc. Printed Name/Title	

Signature	Date
Thomas H. Oliphant, PhD Executive Director, Clinical Statistics, Innovative Analytics Inc. Printed Name/Title	

Signature	Date
Stefanie Senger, PhD Biostatistician, Momentum Research, Inc. Printed Name/Title	

VERSION HISTORY

Version Number	Version Date	Summary and rationale for change(s)
Final 1.0	21-June-2018	N/A
Final 1.1	6-September-2018	Added shift tables for analysis of changes in HbA1c
Final 1.2	21-September-2018	Definition of baseline assessment modified to allow inclusion of a few measures taken minutes after the first dose of study medication
Final 1.3	01-October-2018	Added possible sensitivity analyses for changes in laboratory measures with imputations to assess impact of differential changes in glycemic control medications, and differential missing follow-up, in placebo versus active patients
Final 1.4	08-April-2019	Added analyses of events to be adjudicated (cause of death, occurrence of major adverse cardiovascular and liver events, and severe hypoglycemia) and fractures requiring surgery or hospitalization. Added summary of log-transformed LFTs. Modified visit windows for unscheduled visits to include Visit 8. Modified calculation of compliance based on capsule counts due to missing data. Clarification of statistical modelling of mean change from baseline of fibrosis score, and other minor corrections.
Final 1.5	16-May-2019	Added further sensitivity analyses for primary and secondary efficacy endpoints. Corrected analysis of specialty biomarker Pro-C3. Modified analysis of study drug compliance.

TABLE OF CONTENTS

GLOSSARY OF ABBREVIATIONS

Abbreviation	Term
AE	Adverse event
ALP	Alkaline phosphatase
ALT	Alanine transaminase (SGPT)
ANCOVA	Analysis of covariance
APRI	Aspartate aminotransferase-to-platelet ratio index
AST	Aspartate transaminase (SGOT)
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass Index
BP	Blood Pressure
CI	Confidence Interval
CEC	Clinical Events Committee
CK-18	Cytokeratin-18
CMH	Cochran-Mantel-Haenszel
CRN	Clinical Research Network
CTX	C-terminal telopeptide
DM	Diabetes mellitus
DPP4	Dipeptidyl peptidase-4
ECG	Electrocardiogram
eCRF	Electronic case report form
EDC	Electronic data capture
ELF	Enhanced liver fibrosis score
F1	Fibrosis stage 1
F2	Fibrosis stage 2
F3	Fibrosis stage 3
FFA	Free fatty acids
FIB-4	Fibrosis 4 index for liver fibrosis
GGT	Gamma-glutamyl transpeptidase
GLP1	Glucagon-like peptide-1
HA	Hyaluronic acid
HbA1c	Glycosylated hemoglobin
HBsAg	Hepatitis B surface antigen
HCC	Hepatocellular carcinoma
HCV	Hepatitis C virus
HDL	High-density lipoprotein
HF	Heart failure
HIV	Human Immunodeficiency Virus
HMW	High molecular weight
HOMA-IR	Homeostasis model assessment-insulin resistance
hsCRP	High-sensitivity C-reactive protein
ICF	Informed consent form
I/E	Inclusion and exclusion
INR	International Normalized Ratio

IP	Investigational product
IXRS	Interactive voice/web response system
LDL	Low-density lipoprotein
MACE	Major Adverse Cardiovascular Event
MedDRA	Medical Dictionary for Regulatory Activities
MELD	Model for End-Stage Liver Disease
MI	Myocardial infarction
mITT	Modified Intent-to-treat
NAFLD	Non-alcoholic fatty liver disease
NAS	NAFLD Activity Score
NASH	Nonalcoholic Steatohepatitis
NMR	Nuclear magnetic resonance
PIIINP	Amino-terminal peptide of type III collagen
PNPLA3	Patatin-like phospholipase domain-containing protein 3
PPARγ	Peroxisome proliferator-activated receptor gamma
PPS	Per-protocol Set
Pro-C3	N-terminal type III collagen propeptide
PT	Prothrombin Time – or – Preferred Term
QTcB	Bazett corrected QT interval
QTcF	Fridericia corrected QT interval
RNA	Ribonucleic acid
SAE	Serious Adverse Event
SAF	Safety Analysis Set
SAP	Statistical Analysis Plan
SD	Standard Deviation
SGLT2	Sodium-glucose cotransporter 2
SOC	System Organ Class
T2DM	Type 2 diabetes mellitus
TEAE	Treatment-emergent adverse event
TESAE	Treatment-emergent serious adverse event
TIA	Transient ischemic attack
TIMP-1	Tissue inhibitor of metalloproteinases
USA	Unstable angina
VLDL	Very low-density lipoprotein
WHO	World Health Organization

1 SOURCE DOCUMENTS

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

Document	Date	Version
Protocol	31-July-2017	Amendment 4
eCRF	11-October-2017	2.0

2 PROTOCOL DETAILS

2.1 Study Objectives

To assess the preliminary safety and efficacy of 3 doses of MSDC-0602K (62.5, 125, or 250 mg) as compared to placebo in subjects with non-alcoholic steatohepatitis (NASH).

2.2 Overall Study Design

This is a randomized, double-blinded, placebo-controlled multiple-dose phase II study to evaluate the safety, tolerability and efficacy of three dose levels of MSDC-0602K in subjects with non-alcoholic steatohepatitis (NASH).

Approximately 380 subjects will be randomized at 1:1:1:1 ratio to the following four treatment arms, and the drug is taken orally once per day for 12 months:

- MSDC-0602K 62.5 mg
- MSDC-0602K 125 mg
- MSDC-0602K 250 mg
- Placebo

Randomization will be stratified by:

- Presence of Type 2 diabetes mellitus (T2DM) (Yes/No)
- Use of Vitamin E ≥ 400 IU (Yes/No)
- Fibrosis score (F1 vs. F2/F3)

This study will comprise a maximum 42-day Screening period, a treatment period of 360 days and a 14-day follow-up period. The total duration of the study for each subject will be up to 416 days, including follow-up.

The primary efficacy endpoint will be assessed at 12 months after randomization. Visits to the clinic will be at baseline, 1, 2, 3, 6, 9, and 12 months, with one follow-up visit at study day 374.

2.3 Sample Size and Power

Each active treatment group will be compared with the placebo group at the two-sided 0.05 significance level; 85 subjects per arm provides approximately 80% power to detect a difference in the proportions of subjects achieving the primary efficacy endpoint of 39.5% in any MSDC-0602K group versus 20% response rate in the placebo group. These response rates are similar to effects reported after 18 months daily treatment with 45 mg pioglitazone (58% response) or placebo (17% response) in subjects with pre-diabetes or T2DM (Cusi 2014); after 96 weeks daily treatment with 30 mg pioglitazone (34% response) or placebo (19%) in non-diabetic adults with NASH in the PIVENS study (Sanyal 2010); and in the GOLDEN study, after 52 weeks daily

treatment with 120 mg elafibranor (32% response) or placebo (21%) in the subset of NASH subjects with a baseline NAS ≥ 4 (Ratziu 2016). Assuming 10% of subjects will not complete the 12-month assessment due to early dropout, 380 subjects will be randomized to the study. Sample size calculations were made using nQuery Advisor® (Statistical Solutions Ltd., Boston, MA, USA).

3 EFFICACY AND SAFETY VARIABLES

3.1 Primary Efficacy Endpoint

The primary efficacy endpoint is hepatic histological improvement in NAS (NASH CRN scoring) at month 12. A responder (subjects with hepatic histological improvement) is defined as having all of the following:

- A decrease of at least 2 points in NAS at 12 months. The reduction in NAS must also include at least a 1 point reduction in either ballooning or inflammation.
- No concurrent worsening of fibrosis, defined as no increase in CRN fibrosis score (i.e., an increase of 1 stage or more).

3.2 Secondary Efficacy Endpoints

Secondary efficacy endpoints include the following:

- Proportion of subjects with resolution of NASH with no worsening of fibrosis at 12 months. Resolution of NASH is defined as a ballooning score of 0 and an inflammation score of 0-1. No worsening of fibrosis is defined as no increase in CRN fibrosis score.
- Proportion of subjects with improvement of fibrosis (CRN staging score) by at least 1 stage with no worsening of NASH (defined as no worsening of lobular inflammation or hepatocellular ballooning) at 12 months
- Mean change from baseline in NAFLD activity score (NAS) and each one of its components (steatosis, inflammation and ballooning) at 12 months
- Mean change from baseline in fibrosis score at 12 months

3.3 Exploratory Efficacy Endpoints

Exploratory efficacy endpoints include the following:

- Mean change from baseline relative to placebo at 12 months in liver stiffness as measured using the FibroScan device
- Change from baseline relative to placebo for liver function tests (ALT, AST, ALP, GGT, bilirubin, albumin, PT and INR), normalization of ALT and AST, and changes from baseline relative to placebo for biomarkers and indirect measures of apoptosis and liver fibrosis (CK-18 [M65], Fibrosure, APRI, FIB-4 index, Pro-C3, and ELF score and its individual components TIMP-1, PIIINP, HA) at 6 and 12 months.

- Changes from baseline relative to placebo in circulating inflammatory markers (HMW adiponectin and hsCRP) and markers of bone metabolism (osteocalcin and serum CTX) at 6 and 12 months.
- Changes from baseline relative to placebo in FFA, serum triglycerides, and fasting cholesterol (total, HDL, LDL, VLDL) at 6 and 12 months, and changes from baseline in NMR analysis of lipoprotein particle size at 12 months.
- Changes from baseline relative to placebo in markers of insulin sensitivity: HOMA-IR, HbA1c, fasting plasma glucose, and fasting insulin at 6 and 12 months.
- Changes from baseline relative to placebo in blood pressure (systolic and diastolic), heart rate, weight, and abdominal girth at 6 and 12 months.
- Changes from baseline relative to placebo at 12 months in the central pathologist's general impression of NASH diagnosis: 1) Confirmed, or 2) Cannot be confirmed.

3.4 Safety Variables

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

- The incidence of treatment-emergent adverse events
- Changes from baseline to each time point in clinical laboratory (hematology, chemistry, urinalysis) values (some of which are also evaluated as efficacy parameters), and the occurrence of clinically significant abnormal changes
- Changes from baseline in vital signs (systolic and diastolic blood pressure, heart rate, weight, and abdominal girth) at each visit, which will also be examined as exploratory efficacy endpoints
- Changes from baseline in quantitative electrocardiogram (ECG) parameters (mean heart rate, PR interval, QRS duration, QT interval, QTcB interval, QTcF interval, RR interval)
- Time to first event of death, adjudicated non-fatal myocardial infarction (MI) or unstable angina (USA) hospitalization, adjudicated hospital admission for heart failure (HF), or adjudicated non-fatal ischemic stroke
- Time to first event of death, adjudicated non-fatal MI or USA hospitalization, adjudicated hospital admission for HF, adjudicated non-fatal ischemic stroke, or adjudicated liver event. A liver event consists of ascites, hepatic encephalopathy, variceal hemorrhage, hepatocellular carcinoma (HCC), or liver transplant
- Incidence of adjudicated severe hypoglycemia events
- Incidence of bone fracture requiring surgery or hospitalization

4 PHARMACOKINETIC/PHARMACODYNAMIC VARIABLES

Steady-state trough measurement will be made for MSDC-0602 and MSDC-0597 (a major pharmacologically active metabolite), at Study Visits 2, 4, 5, 6, and 7 (Days 30, 90, 180, 270, 360). The details for the population PK analysis will be covered in a separate Pharmacokinetic Analysis Plan.

5 ANALYSIS POPULATIONS

Final decisions regarding the assignment of subjects to analysis sets will be made during the Validity Review Meeting and documented in the Validity Review Report prior to final database lock and unblinding for the final analysis. **Data from sites with serious GCP or quality concerns may be excluded from the analyses.** The Sponsor will provide a list of such sites, and during the review it will be decided from which analyses and from which sites data should be excluded.

5.1 Enrolled Set

The Enrolled set includes all subjects who signed informed consent.

5.2 Modified Intent-to-treat Analysis Set

The modified Intent-to-Treat (mITT) analysis set includes all subjects who are randomized and who receive at least one dose of study drug. Baseline measures and post-baseline assessments through 14 days after the last dose of study drug will be included in the analyses unless stated otherwise below. Subjects in the mITT will be analyzed as randomized regardless of treatment received. This means that subjects who are randomized who receive the incorrect study drug (i.e., not the study drug to which they were randomized) will be analyzed for efficacy under mITT as in the treatment group to which the subject was randomized. The primary population for efficacy analyses is the mITT.

5.3 Safety Analysis Set

The Safety analysis set (SAF) includes all subjects who are randomized and received at least one dose of study drug. SAF will be used in the analyses of all safety endpoints. Subjects will be included in the analyses according to the actual treatment they received. If a subject incorrectly receives more than one dose of study drug, the subject will be analyzed in the treatment group corresponding to the highest dose received. Baseline measures and post-baseline assessments through 14 days after the last dose of study drug will be included in the analyses.

5.4 Per Protocol Analysis Set

The per-protocol analysis set (PPS) includes all subjects in mITT without major protocol deviations. Major protocol deviations are defined as those deviations considered to have a major effect on the collection or interpretation of the primary efficacy endpoint. Section 5.4.1 details the classification of deviations. Additionally, subjects assigned to the active treatment groups with 2 or more samples in which PK trough levels of both the parent drug (MSDC-0602K) and its metabolite are below the level of detection will be assumed to have been non-compliant and may be excluded from the PPS, and/or all subjects at sites where the majority of subjects assigned to active treatment appear to have been non-compliant based on trough levels may be excluded. Identification of all subjects and measures to be included in the PPS analyses will be determined before the final database lock and unblinding for the final analysis.

The PPS will be used as an additional sensitivity analysis for key efficacy endpoints. PPS subjects will be analyzed as randomized regardless of treatment received.

5.4.1 Important Protocol Deviations Leading to Exclusion from the PPS Analysis

Only those deviations considered to have a major effect on efficacy will lead to complete exclusion of the subjects from the PPS population. For the purposes of this study, the following criteria have been identified as important protocol deviations as it is considered that the occurrences of any of these criteria might have an important influence on the primary efficacy endpoint. Other deviations not specifically identified here may result in exclusion from the PPS analysis and will be identified prior to final database lock and unblinding.

Type	Deviation	Method of Identification
Inclusion and Exclusion (I/E) Criteria	Subjects who fail to comply with key I/E criteria	**Programmatic** check to identify subjects based on protocol deviations and CRF data who failed to meet the following major eligibility criteria: **Exclusion Criteria** (*additions to the exclusion criterion specified in the protocol are marked in italics*): 1. Known history of HIV. 2. Prior liver transplantation. 3. Other well-documented causes of active chronic liver disease, including, but not restricted to: history of positive hepatitis B surface antigen (HBsAg) or HCV RNA (subjects must be viremia free for a minimum of 2 years), suspicion of drug-induced liver disease, alcoholic liver disease and auto-immune hepatitis, Wilson's disease hemochromatosis, primary biliary cirrhosis, primary sclerosing cholangitis, genetic hemochromatosis, known or suspected HCC, history of planned liver transplant or current MELD score > 12. 4. History of cirrhosis and/or hepatic decompensation including ascites, hepatic encephalopathy or variceal bleeding. Platelet count < 100,000/μL of blood at Screening. 5. AST or ALT >5 times the upper limit of normal at Screening. 8. History of alcohol abuse or drug abuse within 6 months of Screening *or discovery of heavy alcohol use during study treatment*. 12. Current use of any diabetes medications other than metformin, GLP-1 agonists, SGLT2 inhibitors, sulfonylureas or DPP4 inhibitors three months prior to the baseline diagnostic liver biopsy or during the study unless as specified in the protocol. Current use of insulin or PPARγ agonists (pioglitazone or rosiglitazone) is not allowed for the duration of the study. 21. Participation in an investigational study (other than a non-treatment registry study) or received an investigational drug within 30 days or 5 half-lives (whichever is longer) prior to *or any time after* randomization. **Inclusion Criteria**: 2. Histological evidence of NASH, based on *the central read of the* biopsy *for purposes of determining eligibility*, with a NAS (NASH CRN scoring) ≥ 4 with a score of at least 1 in each component of NAS. 3. Histological evidence of liver fibrosis defined as NASH CRN System fibrosis score of F1 to F3.

Type	Deviation	Method of Identification
		Thus, subjects will be reviewed for exclusion if NAS < 4, or score of 0 in any component of NAS, or fibrosis score of F0 or F4. 5. Subjects with type 2 diabetes mellitus (DM) must be under stable and reasonable control, defined as under a healthcare provider's care for DM with glycosylated hemoglobin [HbA1c] ≤ 9.5%. Subjects taking permitted anti-diabetic therapies (metformin, GLP-1 agonists, SGLT2 inhibitors, sulfonylureas or DPP4 inhibitors) must be on stable doses for a period of at least 3 months prior to randomization. **Manual** review will be performed for identifying subjects who violate the above I/E criteria, including review of prohibited changes reported for the 3 months prior to randomization (in reference to inclusion criterion #5 above).
Prohibited Medication Use	Subjects who took medications that are not permitted during the study	**Programmatic** identification of subjects with CRF-reported post-randomization use of PPARγ agonist, insulin, ursodeoxycholic acid, paclitaxel, phenytoin, warfarin, celecoxib, tolbutamide or repaglinide; and subjects with reported protocol deviations related to prohibited medication use. **Manual** review of blinded concomitant medications listing to identify subjects with prohibited changes post-randomization in therapeutic agents for diabetes. The medical monitor and Sponsor will review the listing and note any prohibited changes in these glycemic control medications; these subjects may be excluded from the PPS.
Non-compliance During the Study	Subjects who had low study drug exposure and compliance	**Programmatic** check of estimated compliance based on the study drug administration data, and programmatic identification of subjects assigned to an active treatment group with parent drug (MSDC-0602K) or MSDC-0597 (major metabolite) trough levels below the detection threshold at two or more visits. **Manual** review will be performed to identify subjects with compliance less than 80% based on study drug administration data (Section 7.7.2) and/or other compliance issues identified by sponsor.
Errors in Treatment Allocation	Subjects who received a wrong treatment at 1 or more study visits due to packaging or dispensing errors during the study	**Programmatic** check based on unblinded IXRS database after the study is unblinded for the final analysis. The check will be done by comparing the bottle number that IXRS had assigned to the subject/visit against the bottle number actually used.

Type	Deviation	Method of Identification
Other Significant Protocol Deviation	Protocol deviations based on the clinical monitoring.	**Manual** review: This list will be reviewed by the medical monitor and Sponsor and important protocol deviations leading to exclusion from the PPS will be identified.

As defined in the table, the majority of the protocol deviations will be determined programmatically from the data. Those criteria that require clinical interpretation will be reviewed prior to final database lock, following discussion with the medical monitor and Sponsor.

All important protocol deviations occurring during the study will be reviewed and approved by the Sponsor prior to final database lock and unblinding for the final analysis. Should other categories of important protocol deviations, not anticipated at the time of preparing this SAP be identified during the study (and prior to unblinding for the final analysis), they will be provided in a separate document and included in all relevant protocol deviation reviews and approvals.

6 DATA HANDLING

6.1 Time points and Visit Windows

Study Day 1

Study day 1 is defined as the date of the first dose or the date of randomization for subjects who were not administered any dose of study drug.

Study Day

Study days after study day 1 are calculated as (assessment date – study day 1 date) + 1. Study days prior to study day 1 are calculated as (assessment date – study day 1 date). The day prior to study day 1 is study day -1, there is no study day 0.

Scheduled and Unscheduled Visits

For the purpose of statistical analysis, the nominal visit designations as assigned by the investigator will be used. Unscheduled visits will be programmatically assigned to a nominal visit designation according to the following definition of visit windows.

Definition of visit windows to assign unscheduled visits to nominal visit designations

Nominal Visit	Target Day (Study Day)	Parameter measured at all visits	Parameter measured at all visits except Visit 8	Parameter measured at all visits except Visits 2, 3, and 8	Parameter measured at Visits 1, 5, 7 and 8	Parameter measured at Visits 1, 5, and 7	Parameter measured at Visits 1 and 7
1 Baseline	1	≤1	≤1	≤1	≤1	≤1	≤1
2 Month 1	30	2-45	2-45	--	--	--	--
3 Month 2	60	46-75	46-75	--	--	--	--

4 Month 3	90	76-135	76-135	2-135	--	--	--
5 Month 6	180	136-225	136-225	136-225	2-270	2-270	--
6 Month 9	270	226-315	226-315	226-315	--	--	--
7 Month 12	360	316-367	≥316	≥316	271-367	≥271	≥271
8 Follow-up	374	≥368	--	--	≥368	--	--

For all populations, multiple visits for a given parameter with the same nominal visit designation will be dealt with as follows:

- If both scheduled and unscheduled visit data are available, the scheduled visit data will be used for analysis with the exception of baseline, where the baseline value will be defined as the last scheduled or unscheduled value collected on or prior to the date of the first dose of treatment (or randomization for subjects who were not administered any dose of study drug).
- If multiple unscheduled visits occur within the same visit window and no data from a scheduled visit is available for the corresponding nominal visit designation, then the unscheduled visit closest to the corresponding target day will be used in the analysis. If there is a tie, the later unscheduled visit will be used in the analysis.

6.2 Handling of Dropouts or Missing Data

In order to achieve the goal of a well conducted clinical trial according to International Council on Harmonisation Good Clinical Practice, every effort should be made to collect all data (i.e., if a subject misses a scheduled assessment, the site personnel should contact the subject and request him/her to come to the clinic for the visit) (International Conference on Harmonisation, 1998; European Medicines Agency, 2011). However, despite best efforts, it may be inevitable that missing or incomplete data are reported. All missing or partial data will be presented in the subject data listings, as they were recorded on the eCRF.

Missing data will not be imputed in statistical analyses, except as noted below (Section 6.2.1 - 6.2.4).

6.2.1 Missing Dates

For AEs with partial or missing onset or stop dates:

AE stop date will be imputed first as:

- If stop date is completely missing, assume it is ongoing (no imputation);
- For a partial AE stop date: if day is missing, use last day of the month; if both day and month are missing, use December 31st.

Then AE onset date will be imputed as:

- If onset date is completely missing: the first dose date;
- For a partial AE onset date (day is missing, or both day and month are missing):
    - Partial date < the first dose date: last day of the month, or December 31st
    - Partial date = the first dose date: the first dose date
    - Partial date > the first dose date: first day of the month or January 1st

If the imputed AE onset date is after the AE stop date/imputed AE stop date, then the onset date will be set to the AE stop date/imputed AE stop date.

The imputed dates will not be listed. Study day relative to the first dose of study drug associated with missing or partial dates will not be displayed in AE listings.

In the event that a partial date (month/year or year) for concomitant medication is available, this information will be used as follows:

- When both month and year are available – first day of the month will be used for start date and the last day of the month will be used for the stop date.
- When only year is available – January 1st will be used for the start date and December 31st will be used for the stop date.

The imputed dates will only be used to determine whether a concomitant medication will be classified as prior medication or concomitant medication.

6.2.2 NASH CRN Histological Scores

Fibrosis scores are reported by the central histopathologist as 0, 1a, 1b, 1c, 2, 3, 4. Scores of 1a, 1b, and 1c will all be considered as fibrosis stage 1 for analysis. Worsening of fibrosis will be considered as an increase of 1 stage or more. Note that the baseline fibrosis score (F1 vs. F2/F3) to be included as a factor in the logistic regression, risk differences, and ANCOVA models will be the value used to determine the stratum at randomization and will come from the IXRS. The categorical fibrosis score summarized with counts and percentages and as a numeric score in the demographic summary, as well as the analyses of fibrosis score as a numeric endpoint will use those values reported by the central histopathologist as follows:

- In case the liver biopsy was repeated due to a clinical reason as determined by the investigator and approved by Cirius, or because the sample was not readable the biopsy with the latest collection date will be used.
- In case there are multiple samples (accession numbers) associated with a biopsy on a given collection date, the results for the accession number associated with the highest NAS will be used for analysis.
- Results from the central histopathologist's blinded reread of the qualifying screening biopsy utilized to determine eligibility will be used as baseline data.

The NASH CRN histological scores as read by the local pathologist and recorded on the eCRF at Screening (BIOPFIB1) and at Visit 7 (BIOPFIB2) will not be used in any analyses but may be included in subject listings.

6.2.3 Imputation for Primary Endpoint

Sensitivity analyses will be performed on the primary endpoint using two different imputation methods. The sensitivity analyses will include all subjects who were randomized and who received at least one dose of study drug. Two approaches will be used to impute or substitute missing response at 12 months. Post-baseline observations that have been excluded from the primary analysis due to biopsies taken more than 14 days after last dose of study drug will be treated as missing and imputed or substituted as well.

When using multiple imputation, missing baseline and/or post-baseline values needed to determine response at 12 months will be imputed multiple times to account for the uncertainty about the true values to impute. To this end, each of the four categorical variables underlying the response definition of the primary endpoint will be imputed if missing, using the following categories:

- Steatosis: 0, 1, 2, 3
- Lobular inflammation: 0, 1, 2, 3
- Ballooning: 0, 1, 2
- Fibrosis score: 0, 1, 2, 3, 4

Using the SAS® procedure PROC MI 10 imputed data sets will be generated. The regression model will be adjusted for the corresponding baseline values and will further use presence of diabetes mellitus, Vitamin E use ≥ 400 IU, age, and sex as covariates.

An example of the SAS® code to be used is below:

```
proc mi data=<input dataset> out=<output dataset> seed=1234 nimpute=10
      minimum=<…> maximum=<…> round=<…>;
var   <12 months steatosis> <baseline steatosis>
      <12 months lobular inflammation> <baseline lobular inflammation>
      <12 months ballooning> <baseline ballooning>
      <12 months fibrosis score> <baseline fibrosis score>
      <presence of diabetes mellitus> <Vitamin E use ≥ 400 IU> <sex><age at baseline>;
run;
```

The 10 multiple imputed datasets will be used separately in the statistical analysis model and the resulting odds ratios and 95% CIs will be combined via Rubin's rules (using PROC MIANALYZE) to provide a global estimate of the odds ratio with a corresponding 95% CI and p-value.

As a second sensitivity analysis of the primary endpoint, subjects with missing response data at 12 months will be included as non-responders. Similar to above, this sensitivity analysis will include all subjects who were randomized and who received at least one dose of study drug.

6.2.4 Imputation for Missing Changes in Laboratory Values

HbA1c

Additional sensitivity analyses may be performed conditional on the distribution of changes in glycemic control medications between the placebo and active treatment groups at 6 and 12 months.

If more than 35% of the subjects who had a major up-titration of glycemic control medications are in the placebo group, a sensitivity analysis may be done where the 6- and 12-month values of HbA1c for subjects who had a major up-titration of glycemic control medications will be imputed to be 0.5 % higher than the observed value.

If more than 15% of the subjects who had a major down-titration of glycemic control medications are in the placebo group, a sensitivity analysis may be done where the 6- and 12-month values of HbA1c for subjects who had a major down-titration of glycemic control medications will be imputed to be 0.5 % lower than the observed value.

If more than 35% of the subjects missing paired change-from-baseline values of HbA1c (separately at 6 and 12 months) are in the placebo group, a sensitivity analysis may be performed where the missing follow-up value will be assigned the baseline value plus 0.5%.

Liver function tests and other markers

Additional sensitivity analyses may be performed conditional on the distribution of missing values between the placebo and active treatment groups at 6 and 12 months.

If > 35% of the subjects missing paired change-from-baseline values of the following laboratory tests (separately at 6 and 12 months) are in the placebo group, a sensitivity analysis may be performed where missing follow-up values will be assigned the corresponding baseline values plus 0.5 times the maximum change (maximum increase) from baseline for the parameter observed in any subject (i.e., regardless of treatment group) at the timepoint: ALT, AST, ALP, Total Bilirubin, GGT, pro thrombin time, APRI score, CK-18, ELF, FIB-4, Pro-C3, Fibro-test score, and HOMA-IR. Likewise for albumin, missing follow-up values may be assigned the corresponding baseline values minus 0.5 times the minimum change (maximum decrease) from baseline .

7 STATISTICAL METHODS

7.1 General Principles

All data processing, summarization and analyses will be performed using Version 9.3 (or later) of the SAS® (SAS Institute, Cary, NC) statistical software package.

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

- Placebo
- MSDC-0602K 62.5 mg
- MSDC-0602K 125 mg
- MSDC-0602K 250 mg
- Total (if applicable)

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

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

Descriptive summary statistics for continuous variables will include the number of observations (N), mean, standard deviation (SD), median, 25th and 75th percentiles, minimum and maximum.

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

Dates will be displayed as DDMMMYYYY.

7.2 Subject Disposition and Data Sets Analyzed

The number and percentage of subjects achieving some study milestones listed below will be summarized by treatment group and overall (if applicable):

- Subjects randomized by site
- Subjects randomized by randomization strata [IXRS data]
- Number and percentage of subjects screened, randomized, treated, included in each study population (mITT, SAF, PPS)
- Number and percentage of subjects who complete the study and who discontinue early, including a breakdown of the primary reasons for discontinuation
- Number and percentage of subjects who complete the administration of study drug, discontinue the study drug early, and breakdown of the primary reasons for early discontinuation of study drug

Listings will be provided for screen failure and discontinued subjects.

7.3 Protocol Deviations

The number and frequency of subjects with one or more protocol deviations identified through clinical monitoring within each protocol deviation category (failure to adhere to visit schedule, assessment not performed per protocol, IP non-compliance, use of prohibited medication, IP dispensing error, ICF issue, Inclusion/exclusion criteria (eligibility), other) will be summarized by treatment group in the mITT population. Additionally, important protocol deviations leading to exclusion from the PPS population (see Section 5.4.1) will be listed and summarized by treatment group for the mITT population.

The important protocol deviations will be identified prior to the final database lock and prior to unblinding for the final analysis.

7.4 Demographics and Other Baseline Characteristics

Demographic and baseline characteristics will be listed and summarized by treatment group and overall for the mITT population. Standard descriptive statistics will be presented for the continuous variables of:

- Age (years)

- Weight (kg)
- Height (cm)
- Body mass index (BMI) (kg/m^2)
- Waist circumference (cm)
- NAFLD activity score (NAS)
- Fibrosis score

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

- Age category (<55 vs. ≥55 years)
- Sex (male vs. female)
- Race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Pacific Islander, White, Other)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino, Not Stated, Unknown)
- Vitamin E ≥ 400 IU (Yes vs. No) [EDC data]
- Fibrosis score (0, 1, 2, 3, 4)
- Presence of T2DM (Yes vs. No) [EDC data]
- Smoking (Never, Current, Former)
- Alcohol abuse (Abstinent, Light, Moderate, Heavy)

Fibrosis score is the score reported by the central histopathologist (see Section 6.2.2).

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

7.5 Medical History

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

7.6 Prior and Concomitant Medications

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

Prior medications and concomitant medications are defined as follows:

- Prior medications are those taken with a start date prior to the first dose date of study treatment
- Concomitant medications are those with a start date on or after the first dose date of study treatment, or those with a start date before the first dose date of study treatment and a

stop date on or after the first dose date of study treatment (including those medications which are classified as ongoing).

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

Prior medications, prohibited medications (refer to protocol sections 6.4 and 7.8) and concomitant medications taken during the study will be summarized for the mITT population. The number and percentage of subjects using each medication will be summarized by ATC-Level 2 class and preferred term by treatment group.

Prohibited medications include PPARγ agonists, insulin, ursodeoxycholic acid, paclitaxel, phenytoin, warfarin, celecoxib, tolbutamide and repaglinide. A disallowed change post-randomization in dose of metformin, GLP1 agonists, SGLT2 inhibitors, sulfonylureas, DPP4 inhibitors, or insulin will be included in the summary of prohibited concomitant medications.

The number and percentage of diabetic subjects (subjects with T2DM) with an up-titration, down-titration, and no major change in glycemic control medications after randomization will also be summarized by treatment group. The classification of each subject into one of the three categories will be based on blinded adjudication based on a manual review of subjects' use of glycemic control medications prior to final database lock and unblinding. The "no major change" category may include subjects with changes identified as not being clinically significant during blinded adjudication.

In addition, all prior and/or concomitant medications will be presented in subject data listings, with additional listings presenting only prohibited medications and only glycemic control medications.

7.7 Treatment Compliance and Exposure

7.7.1 Exposure

Overall treatment duration in days will be evaluated and summarized from the first dose date to the last dose date (as recorded on the End of Treatment CRF page) of randomized study drug using the following formula:

Overall treatment duration = end of treatment date – first dose date + 1.

Descriptive statistics for overall treatment duration (days) will be presented by treatment group and will also be summarized by duration categories:

- < 90 days
- ≥90 and < 180 days
- ≥180 and < 270 days
- ≥270 and <360 days
- ≥360 days

7.7.2 Study Drug Administration Data

Study drug capsule count will be conducted at Visit 4 (3 months), Visit 5 (6 months), Visit 6 (9 Months), and Visit 7 (12 months), and possibly at Visit 2 (1 month), Visit 3 (2 Months), and Visit 8 (Follow-up). Overall compliance will be calculated as the actual number of capsules

taken divided by the number of capsules required, then multiplied by 100. The required number of capsules will be calculated as the number of capsules per day (1) times the overall duration of treatment administration in days. The administration duration in days will be derived from the drug dispense date in the drug administration log as:

Administration duration = last dose date – first dispense date + 1.

The last dose date will be the recorded end of treatment date if available. If the end of treatment date is missing, the end of study date will be used instead if available.

The following special cases may apply per subject:

- If the return capsule count is missing for a drug dispensation record other than the last, the duration of the corresponding period will be subtracted from the overall administration duration. The duration to be subtracted will be derived as return date minus dispense date of that record. If the return date is missing as well, the dispense date of the following record will be used as end date.
- If the return capsule count is missing for the last drug dispensation record (e.g. patient was lost to follow-up), the administration duration will be calculated using the return date of the latest available record with non-missing dispense and return counts as last dose date.
- If only the return date is missing for any record but dispense count, return count, and dispense date are available, the dispense date of the following record will be used as substitute return date if available. Otherwise the last dose date as defined above will be used.
- If study drug was dispensed after last dose date by mistake, the corresponding records will not contribute to the overall compliance result.

As a summary, the following formula will be used to calculate overall compliance across records with available data and after taking special cases as described above into account:

$$100 \times \frac{(total\ no.of\ capsules\ dispensed - total\ no.of\ capsules\ returned)}{administration\ duration\ in\ days}$$

Subjects with an overall compliance less than 80% will be reviewed by the medical monitor and Cirius Therapeutics, Inc., for potential exclusion from the per-protocol analyses. Exclusion of non-compliant subjects from the per-protocol analysis set will be pre-specified prior to the final database lock and prior to unblinding for the final analysis.

7.7.3 Pharmacokinetic Trough Levels

Trough levels of parent drug (MSDC-0602K) and its metabolite (MSDC-0597) will be listed, with subjects assigned to active treatment identified as potentially non-compliant based on these levels.

7.8 Efficacy

7.8.1 Primary Efficacy Analysis

The primary efficacy endpoint is hepatic histological improvement in NAS (NASH CRN scoring). A responder (subjects with hepatic histological improvement) is defined as having all of following:

- A decrease of at least 2 points in NAS at 12 months. The reduction in NAS must also include at least a 1 point reduction in either ballooning or inflammation.
- No worsening of fibrosis, defined as no increase in CRN fibrosis score of ≥ 1 stage.

Each MSDC-0602K dose group will be compared with the placebo group at the two-sided 0.05 significance level. For each comparison, the null hypothesis is that there is no difference in response rates between the MSDC-0602K dose and placebo. The alternative hypothesis is that there is a difference in response rates between the MSDC-0602K dose and placebo. The primary analysis of the primary efficacy endpoint will be performed in the mITT population, with supportive analysis in the PPS population. The number and percentage of responders will be summarized by treatment group. Analyses of liver biopsy findings will be based on the central expert histopathologist's repeat read of the Screening biopsy and read of the Visit 7 biopsy. The central read of the Screening biopsy for purposes of determining subject eligibility, and the local pathologist's assessments, will not be considered for these analyses. Covariate adjustment for randomization stratification factors diabetes mellitus, baseline fibrosis (F1 versus F2/F3), and vitamin E use will utilize values reported at randomization and which determined the stratum in which the subject was randomized.

The primary analysis will be conducted using observed data included in the analysis dataset. Five additional sensitivity analyses will be conducted in the mITT population where (1) missing values are multiply imputed, (2) subjects with missing data will be treated as non-responders, as outlined in Section 6.2.3, (3) subjects with biopsy results up to and including (≤) 1 day after last dose of study drug will be included, (4) subjects with biopsy results up to (<) 7 days after last dose of study drug will be included, and (5) all observed data will be used (including biopsy results obtained more than 14 days after last dose of study drug). These sensitivity analyses will include all subjects who were randomized and who received at least one dose of study drug.

Comparisons between each active treatment group and the placebo group will be evaluated using appropriate contrasts from a logistic regression model including treatment and adjusted for presence of diabetes mellitus, baseline fibrosis F1 versus F2/F3, and ordinal baseline NAS. Model-adjusted odds ratios for each active dose group relative to placebo and their associated 95% confidence intervals (CIs) and p-values will be provided.

In addition, stratification-adjusted differences in response rates (risk differences) adjusted for presence of diabetes mellitus, baseline fibrosis F1 versus F2/F3, and ordinal baseline NAS between each active dose group and placebo and their associated 95% CIs and p-values will be provided (LaVange et al 2005; Zink and Koch 2012).

Supportive analyses of the primary endpoint will include comparisons of each active dose group with placebo using Cochran-Mantel-Haenszel (CMH) tests stratified by presence of diabetes mellitus.

Sensitivity analysis for the primary endpoint will be performed using a logistic regression model including treatment and adjusting for presence of diabetes mellitus, baseline fibrosis F1 versus F2/F3, sex, age, and ordinal baseline NAS. Adjusted odds ratios relative to placebo and the corresponding 95% CIs and p-values will be provided.

A summary of the planned primary efficacy analyses is given below:

Method	**mITT Observed data**	**PPS Observed data, data exclusions possible due to non-compliance**	**mITT Multiply imputed response data**	**mITT Missing response status at 12 months set to non-responder**	**mITT Observed data ≤1 day after last dose of study treatment**	**mITT Observed data <7 days after last dose of study treatment**	**mITT Observed data including results any time after last dose of study treatment**
Logistic regression (odds ratios) including treatment group and adjusted for presence of diabetes mellitus, baseline fibrosis F1 versus F2/F3, ordinal baseline NAS.	X*	X	X	X	X	X	X
Stratification-adjusted differences in response rates (risk differences) between each active dose group and placebo adjusted for presence of diabetes mellitus, baseline fibrosis F1 versus F2/F3, ordinal baseline NAS.	X	X	X	X	X	X	X
CMH tests of each active dose group versus placebo stratified by presence of diabetes mellitus.	X	X					
Logistic regression (odds ratios) including treatment group and adjusted for presence of diabetes mellitus, baseline fibrosis F1 versus F2/F3, sex, age, ordinal baseline NAS.	X	X					

*Primary analysis.

7.8.2 Secondary Efficacy Analysis

Primary analyses of the secondary efficacy endpoints will be conducted in the mITT population with supportive analyses in the PPS population. Descriptive statistics for secondary efficacy endpoints will be presented by treatment group and by study visit where appropriate.

Proportion of subjects with resolution of NASH at month 12

Resolution of NASH is defined as a ballooning score of 0 and an inflammation score of 0-1 without worsening of fibrosis at month 12. No worsening of fibrosis is defined as no increase in CRN fibrosis score of ≥ 1 stage.

The incidence rate of resolution of NASH will be summarized by treatment group. Each active treatment group will be compared to the placebo group using appropriate contrasts from a logistic regression model including treatment and adjusted for presence of diabetes mellitus, baseline fibrosis F1 versus F2/F3, and ordinal baseline NAS. Model-adjusted odds ratios for each active dose group relative to placebo and their associated 95% CIs and p-values will be provided. In addition, stratification-adjusted rate differences and their associated 95% CIs and p-values will be provided, with adjustment for presence of diabetes mellitus, baseline fibrosis F1 versus F2/F3, and ordinal baseline NAS. Supportive analyses will include comparisons of each active dose group with placebo using Cochran-Mantel-Haenszel (CMH) tests stratified by presence of diabetes mellitus. A logistic regression model adjusting for treatment, presence of diabetes mellitus, baseline fibrosis F1 versus F2/F3, sex, age, and ordinal baseline NAS will be used as sensitivity analysis. Adjusted odds ratios relative to placebo and the corresponding 95% CIs and p-values will be provided. Additional sensitivity analyses will be conducted in the mITT population where (1) subjects with biopsy results up to and including (≤) 1 day after last dose of study drug will be included, (2) subjects with biopsy results up to (<) 7 days after last dose of study drug will be included, and (3) all observed data will be used (including biopsy results obtained more than 14 days after last dose of study drug). These additional sensitivity analyses will be applied to the logistic regression model including treatment and adjusted for presence of diabetes mellitus, baseline fibrosis F1 versus F2/F3, and ordinal baseline NAS, and to the analysis of stratification-adjusted rate differences with adjustment for presence of diabetes mellitus, baseline fibrosis F1 versus F2/F3, and ordinal baseline NAS.

Proportion of subjects with improvement of fibrosis by at least 1 stage with no worsening of NASH at 12 months

Reduction of fibrosis is measured by CRN staging score. All subjects with CRN staging score ≥ 1 decrease from baseline will be considered as responders. Worsening of NASH is defined as an increase of either ballooning or inflammation.

The incidence rate of the responders will be summarized by treatment group. Each active treatment group will be compared to the placebo group using appropriate contrasts from a logistic regression model including treatment and adjusted for presence of diabetes mellitus, baseline fibrosis F1 versus F2/F3, and ordinal baseline NAS. Adjusted odds ratios relative to placebo and their associated 95% CIs and p-values will be provided. In addition, stratification-adjusted risk differences and their associated 95% CIs and p-values will be provided, with adjustment for presence of diabetes mellitus, baseline fibrosis F1 versus F2/F3, and ordinal baseline NAS. A logistic regression model adjusting for treatment, presence of diabetes mellitus, baseline fibrosis F1 versus ≥ F2/F3, sex, age, and ordinal baseline NAS will be used as

sensitivity analysis. Adjusted odds ratios relative to placebo and the corresponding 95% CIs and p-values will be provided. Additional sensitivity analyses will be conducted in the mITT population where (1) subjects with biopsy results up to and including (≤) 1 day after last dose of study drug will be included, (2) subjects with biopsy results up to (<) 7 days after last dose of study drug will be included, and (3) all observed data will be used (including biopsy results obtained more than 14 days after last dose of study drug). These additional sensitivity analyses will be applied to the logistic regression model including treatment and adjusted for presence of diabetes mellitus, baseline fibrosis F1 versus F2/F3, and ordinal baseline NAS, and to the analysis of stratification-adjusted rate differences with adjustment for presence of diabetes mellitus, baseline fibrosis F1 versus F2/F3, and ordinal baseline NAS.

Mean change from baseline in NAS and each one of its components (steatosis, inflammation and ballooning) at 12 months

For each outcome separately, the change from baseline to month 12 will be analyzed using an Analysis of Covariance (ANCOVA) model with fixed terms for treatment, age, sex, presence of diabetes mellitus, Vitamin E ≥ 400 IU, baseline fibrosis F1 versus F2/F3, and either ordinal baseline NAS or the baseline value of the component score as applicable (and used as an ordinal covariate). Least square mean treatment differences relative to placebo and associated 95% CIs and p-values will be presented. Summary statistics for baseline values, 12-months values, and change from baseline will be presented for each treatment group.

Mean change from baseline in fibrosis score at 12 months

Mean change from baseline in fibrosis score will be analyzed similarly as the mean change from baseline in NAS above using an ANCOVA model with fixed terms for treatment, age, sex; randomization stratification factors presence of diabetes mellitus, Vitamin E ≥ 400 IU, baseline fibrosis F1 versus F2/F3; and baseline fibrosis score from the repeat read of the screening biopsy (used as an ordinal covariate with possible values 0-4).

7.8.3 Subgroup Analysis

Subgroup analyses of the primary endpoint and secondary efficacy endpoints will be performed for different categories of baseline characteristics and the randomization stratification factors:

- Age at study entry (< 55 years vs. ≥ 55 years)
- Sex (Male vs. Female)
- Race (White vs. Non-white)
- Presence of diabetes mellitus (Yes vs. No)
- Use of Vitamin E ≥ 400 IU (Yes vs. No)
- Fibrosis scores (F1 vs. F2/F3)
- Baseline NAS (≤ 4 vs. > 4)
- PNPLA3 genotype variant (cc/cg vs. gg)

For each dichotomous endpoint, the homogeneity of the effect of each active dose relative to placebo across each subgroup will be assessed using appropriate contrasts in a logistic regression model that includes terms for the treatment groups, the subgroup under consideration, and their interaction, together with adjustment for presence of diabetes mellitus, baseline fibrosis F1 versus F2/F3, and ordinal baseline NAS or the baseline component score, as appropriate. For

quantitative outcomes, similar analyses using ANCOVA will be performed. All subgroup analyses will be exploratory and performed for the mITT population only.

7.8.4 Exploratory Analysis

Exploratory efficacy endpoints are listed in Section 3.3.

The analysis of mean change from baseline at 12 months in liver stiffness will be restricted to subjects for whom liver stiffness has been measured using the FibroScan device. The type of device for each subject and time-point was not recorded in the eCRF but will be captured in an external spreadsheet and provided by the Sponsor prior to unblinding for the final analysis.

All exploratory analysis will be performed for the mITT population. For change from baseline endpoints, baseline values, post-baseline values, and change from baseline will be summarized by treatment groups at each evaluation time point. Highly skewed variables will be log-transformed for analysis, and geometric mean changes presented. Comparisons between each active dose and placebo will be done using ANCOVA models for continuous variables, similarly as for secondary endpoints (refer to Section 7.8.2). Regardless of skewness, summaries and ANCOVA results will be presented for log-transformed values of ALT, AST, ALP, and GGT, in addition to results based on non-transformed values.

Fibrosure numeric results (rather than results expressed as fibrosis stage) will be analyzed. Changes in diagnostic category (NASH diagnosis confirmed or cannot be confirmed) from baseline to month 12 will be presented by means of a shift table.

Changes in markers of insulin sensitivity (HOMA-IR, HbA1c, fasting plasma glucose, and fasting insulin) will be analyzed overall and then separately in subjects with T2DM, and in subjects with T2DM without a major up- or down-titration in glycemic control medications during the study (categorized into the “no major change” category during blinded adjudication).

Mean change from baseline with standard error bars will be plotted for each treatment group by visit for insulin sensitivity markers (HOMA-IR, HbA1c, fasting plasma glucose, and fasting insulin) in the mITT population for (1) subjects with T2DM, and (2) subjects with T2DM without major changes in their glycemic control medications during the study.

Normalization of ALT and AST will be evaluated with shift tables presenting movement in and out of reference range from baseline to each scheduled post-baseline visit as defined in Section 7.9.2 and will be provided for each treatment group in the mITT population.

Changes in HbA1c will also be evaluated with shift tables presenting movement in and out of reference ranges from baseline to 6 and to 12 months. These shift tables will be presented in (1) subjects without T2DM, (2) subjects with T2DM, and (3) subjects with T2DM without major changes in glycemic control medications. The HbA1c “normal range” will be considered 4.0-5.6% in subjects without T2DM and 5.7-6.4% in subjects with T2DM.

7.9 Safety

7.9.1 Adverse Events

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

- TEAEs are events with start date on or after the date of first dose of study treatment and up to 14 days after date of last dose of study treatment or events with start date prior to the date of first dose of study treatment whose severity worsens on or after the date of first dose of study treatment.

Listings of Serious AEs (SAEs), AEs leading to discontinuation of study treatment, and AEs resulting in death will be produced by treatment group. Adjudicated cause of death will be described. AEs with an onset between signing of informed consent and study drug initiation, and those with onset more than 14 days after the last dose of study drug will be listed separately.

The number and percentage of subjects reporting each TEAE will be summarized by System Organ Class (SOC) and Preferred Term (PT) by treatment group and overall in the SAF population. Each subject will be counted only once within each PT or SOC and recording period. If a subject experiences more than one AE within a PT for the same recording period, only the AE with the strongest relationship or the worst severity, as appropriate, will be included in the summaries of relationship and severity. AEs with missing classifications regarding relationship to study drug and start date greater or equal to start of study drug will be considered as related to study drug. AEs that were reported to be at least possibly related to study drug will be considered as related to study drug.

Tables will be produced for the following TEAEs overall and by SOC and PT, preceded by an overview table:

- Any TEAE
- Most common TEAEs (≥ 5% in any treatment group)
- TEAEs by maximum severity
- TEAE related to study drug
- TEAEs related to study drug by maximum severity
- TEAE leading to study drug discontinuation
- TEAEs related to study drug and leading to study drug discontinuation
- TEAEs leading to study drug interruption
- Any TESAE (treatment-emergent serious AE)
- TESAE related to study drug
- TESAE leading to study drug discontinuation
- TESAEs related to study drug and leading to study drug discontinuation
- TESAEs leading to study drug interruption

- TESAE leading to death

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

7.9.2 Laboratory Evaluations

Values for the following safety laboratory tests (hematology, serum chemistry, and urinalysis) will be listed and summarized by treatment group and visit.

Hematology	Serum Chemistry	Urinalysis
Hemoglobin	Albumin*	Bilirubin
Hematocrit	Alkaline phosphatase (ALP)*	Blood
Platelets, RBC with indices	Alanine aminotranferase (ALT)*	Glucose
WBC with differential	Aspartate Aminotransferase (AST)*	Ketones
HbA1c*	Blood Urea Nitrogen (BUN)	pH
INR*	Bicarbonate	Creatinine
PT*	Calcium	Protein
	Chloride	Specific gravity
	Creatinine	Microscopic exam (if blood or protein positive)
	Fasting plasma glucose (FPG)*	
	Fasting insulin*	
	Free fatty acid (FFA)*	
	Lactate dehydrogenase (LDH)	
	Phosphorus	
	Potassium	
	Sodium	
	Gamma-Glutamyl transpeptidase (GGT)*	
	Total Bilirubin*	
	Total protein	
	Uric acid	
	Cholesterol (total, HDL, LDL, VLDL)*	
	Triglycerides*	

*Also considered an efficacy parameter.

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

For analysis purposes, values preceded by a "<" or a ">" sign (i.e., those below or above the limits of quantification) will be considered equal to 0.5 × the lower limit or 1.5 × the upper limit of quantification, respectively. The proportion of subjects with values below the lower limit and above the upper limits of quantification will be provided among the summary statistics.

Laboratory evaluations measured on a continuous scale will be summarized by visit using standard descriptive statistics for the SAF population. Changes from baseline will also be summarized. Summaries will be presented for log-transformed values of ALT, AST, ALP, and GGT, in addition to summaries of non-transformed values. For analysis by visit, analysis windowing as described in Section 6.1 will be used for each scheduled visit such that unscheduled visits will also be considered. For each laboratory analyte, the baseline value will be defined as last scheduled or unscheduled value collected prior to the first dose of treatment.

For hematology and serum chemistry, shift tables presenting movement in and out of reference range from baseline to each scheduled post-baseline visit will be provided for each treatment group. An additional shift table will be produced for ALT where the upper normal limit is defined as 19 U/L for women and 30 U/L for men. Corresponding shift tables (normal vs. high /

normal vs. abnormal) will be produced for urinalysis. Additional tables will summarize the occurrence at each visit of clinically significant abnormal changes from baseline by treatment group for the following analytes: ALT, AST, GGT, and alkaline phosphatase with abnormal change defined as ratio to baseline ≥ 5, and total bilirubin with abnormal change defined as ratio to baseline ≥ 2.

Mean change from baseline with standard error bars will be plotted for each treatment group by visit for select laboratory parameters (hemoglobin, ALT, ALP, AST, Creatinine, total bilirubin) for the SAF population.

7.9.3 Vital Signs

The following vital signs will be summarized by treatment group and visit.

- Sitting systolic and diastolic blood pressure (mm Hg)*
- Heart rate (bpm)*
- Weight (kg)*
- Waist circumference (cm)*

*Also evaluated as an efficacy parameter.

Vital signs data and changes from baseline in vital signs will be summarized by visit using standard descriptive statistics for the SAF population. The baseline value will be defined as last scheduled or unscheduled value collected prior to the first dose of treatment. For analysis by visit, analysis windowing as described in Section 6.1 will be utilized for each scheduled visit such that unscheduled visits will also be considered.

Mean change in weight from baseline with standard error bars will be plotted for each treatment group by visit in the SAF population.

7.9.4 Electrocardiograms

Electrocardiogram (ECG) measurements and changes from baseline in ECG measurements will be summarized by visit using standard descriptive statistics for the SAF population. If a subject has both central and local ECGs for a given time point, the local values will be used for analysis. Summary statistics for the following parameters will be presented: ECG mean heart rate, PR interval, QRS duration, QT interval, QTcB interval, QTcF interval, RR interval. Number and percentage of subjects with atrial fibrillation, sinus rhythm, or other rhythm will be presented by visit.

7.9.5 Physical Examination

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

For each physical examination body system, the number and percentage of subjects with abnormalities at baseline and post-baseline will be summarized by treatment group for the SAF population.

The distribution of edema grades reported at baseline and at each post-baseline visit will be presented by treatment group.

7.9.6 Major Adverse Cardiovascular Events and Liver Events

A Clinical Events Committee (CEC) will adjudicate Major Adverse Cardiovascular Events (MACE), which are defined as death, adjudicated non-fatal MI or USA hospitalization, adjudicated hospital admission for HF, or adjudicated non-fatal ischemic stroke. Time from baseline to occurrence of first event will be computed in days. Subjects without an event will be censored at the earlier of the end of the time period of interest and the last contact date.

Time to first MACE through 14 days after last dose of study treatment will be presented as number of events with Kaplan-Meier estimates of the cumulative risk of event. Hazard ratios and corresponding 95% confidence intervals comparing each active treatment group to placebo will be estimated from a Cox proportional hazards model including a factor for treatment group; and the model will be stratified by randomization stratification factors (presence of diabetes mellitus, Vitamin E $\geq$ 400 IU, and baseline fibrosis F1 versus F2/F3) if the number of events permits. Treatment groups will be compared using a log-rank test.

Similarly, time to first MACE or adjudicated liver event (ascites, hepatic encephalopathy, variceal hemorrhage, hepatocellular carcinoma, or liver transplant) through 14 days after last dose of study treatment will be analyzed and presented.

Results will also be provided for each component of these composite events. For non-fatal events, subjects who die before the time point of interest and who did not reach the event of interest before death will be censored on the day of death. Results will be presented through 14 days after last dose of study treatment for:

- Time to death
- Time to first adjudicated non-fatal MI
- Time to first adjudicated USA hospitalization
- Time to first adjudicated hospital admission for HF
- Time to first adjudicated non-fatal ischemic stroke
- Time to first adjudicated liver event

In addition, deaths will be presented in a listing including investigator-reported data and the adjudicated primary cause of death.

7.9.7 Severe Hypoglycemic Events

The CEC will also adjudicate severe hypoglycemia events. The total number of events and the number and percentage of subjects with any event through 14 days after last dose of study treatment will be presented by treatment group, as adjudicated by the CEC.

7.9.8 Bone Fractures Requiring Surgery or Hospitalization

The number of fractures requiring surgery or hospitalization as well as the number and percentage of subjects with any fracture requiring surgery or hospitalization, as reported by the investigator, through 14 days after last dose of study treatment will be presented by treatment group.

7.10 Interim Analysis

An interim analysis of exploratory and safety endpoints is planned after all subjects have been randomized into the study and 6-month data are available for a minimum of 300 randomized subjects. Changes in selected serum biomarkers from baseline to 6 months will be assessed. Descriptive summary statistics by treatment group will be presented and results shared with a limited group of people to be identified by the Sponsor. The treatment assignments of individual subjects will not be presented, i.e., data listings will not be included for this interim analysis. Details will be provided in a separate SAP for the interim analysis.

8 CHANGES FROM PROTOCOL-SPECIFIED ANALYSES

The specialty biomarkers to be analyzed were changed following finalization of the protocol. The following biomarker described in the protocol will not be analyzed: CK-18 [M30]. Changes from baseline in biomarkers and indirect measures of apoptosis and liver fibrosis at 6 months were added as pre-specified exploratory efficacy endpoints.

The exploratory endpoint "Changes from baseline relative to placebo at 12 months in the pathologist's general impression" has been corrected to reporting the general impression as provided by the central histopathologist: NASH diagnosis 1) confirmed, or 2) cannot be confirmed. While the local pathologists provide the information as initially planned in the protocol, this is not available from the central histopathologist.

The last exploratory endpoint "Changes from baseline relative to placebo in subjects with PNPLA3 variants and other genotypes" has been removed from the list as it is not an endpoint per se but refers to subgroup analyses of efficacy endpoints with respect to genetic variants (see Section 7.8.3).

The definition of the per protocol analysis set was expanded to allow possible exclusion of subjects based on trough levels of study drug and its metabolite.

For the derivation of treatment compliance, the last dose date will be used instead of last dose date +14 days to define the end of the treatment period if study treatment was discontinued early. The derivation of compliance measures per treatment period as well as production of listings and/or presentation of summary statistics of per-period and of overall compliance will be omitted. The rationale for these changes is that because capsule counts are known to be unreliable and despite our best efforts to clean the data there are still discrepancies that cannot be resolved.

The methods for efficacy analysis have been clarified. Due to the expected highly unequal marginal distribution of the stratification factor "use of Vitamin E $\geq$ 400 IU" (Yes vs. No) this factor was removed from all logistic regression models (except for the corresponding subgroup analysis).

The pairwise comparisons of active treatment versus placebo through CMH tests have been revised to only include stratification for presence of diabetes mellitus as to avoid overly sparse cross-classification frequencies of the originally included stratification factors, treatment, and outcome.

According to the protocol, sensitivity analyses had been planned for the primary and secondary dichotomous efficacy endpoints examining the addition of sex to the primary logistic regression

models. The sensitivity analyses have been updated to include age at baseline as well because age is expected to be associated with response, and to exclude use of Vitamin E ≥ 400 IU similar to the primary analysis. Further sensitivity analyses have been added to examine different time restrictions for the biopsies taken at 12 months.

Adjudication of Major Adverse Cardiovascular Events, liver events, and severe hypoglycemic events by a Clinical Events Committee, and collection of information regarding fractures were introduced after finalization of the protocol. Corresponding analyses have been added to the SAP.

9 REFERENCES

Cusi, K. Management of NAFLD/NASH in Patients with T2DM: New Concepts and Future Directions. Presented at American Diabetes Association Annual Meeting, San Francisco, California, June 13-17, 2014.

European Medicines Agency Committee for Medicinal Products for Human Use. Guideline on Missing Data in Confirmatory Clinical Trials. European Medicines Agency: London, 2011. Available at http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. Statistical Principles for Clinical Trials, Guideline E9, 1998. Available at http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf

LaVange LM, TA Durham, and GC Koch. Randomization-based nonparametric methods for the analysis of multicenter trials. Statistical Methods in Medical Research (2005); 14:281-301.

Ratziu V, Harrison SA, Francque S, Bedossa P, Lehert P, Serfaty L, et al. Elafibranor, an agonist of the peroxisome proliferator-activated receptor-α and -δ, induces resolution of nonalcoholic steatohepatitis without fibrosis worsening. Gastroenterology (2016) May;150(5):1147-1159.

Sanyal AJ, Chalasani N, Kowdley KV, McCullough A, Diehl AM, Bass NM, et al. Pioglitazone, vitamin E, or placebo for nonalcoholic steatohepatitis. N Engl J Med (2010) 362:1675-1685.

Zink RC and Koch G. NParCov3: A SAS/IML Macro for Nonparametric Randomization-Base Analysis of Covariance. Journal of Statistical Software (2012); July 50(3):1-17.

10 APPENDIX

Schedule of Activities

	Screen[1]	Treatment							Follow-up
Visit Number		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Study Months		**Rando-mization**	**1**	**2**	**3**	**6**	**9**	**12**	
Study Day	**-42 to -1**	**1**	**30**	**60**	**90**	**180**	**270**	**360**	**374**
Window		**N/A**	**± 7 days**						**± 3 days**
General Activities									
Informed Consent	X								
Medical History	X								
Complete Physical Examination[2]	X								X
Limited Physical Examination[2]		X	X	X	X	X	X	X	
Vital signs[3]	X	X	X	X	X	X	X	X	X
12-lead ECG	X					X		X	X
Pregnancy test (females only)	X	X							X
FSH (post-menopausal females only)	X								
FibroScan[4]	X							X	
Liver Biopsy[5]	X							X	
Randomization		X							
Dietary guidance and assessment of alcohol intake[6]	X	X	X	X	X	X	X	X	
Adverse event review	X	X	X	X	X	X	X	X	X
Concomitant medication review	X	X	X	X	X	X	X	X	
Monthly phone assessment[7]					X	X	X		
Study drug dispensing[8]		X			X	X	X		
Central Lab Evaluations									
Hematology & serum chemistry [9, 10]	X	X	X	X	X	X	X	X	X
Liver function tests	X	X	X	X	X	X	X	X	
Urinalysis	X	X			X	X	X	X	X
HMW adiponectin, hsCRP, CK-18 [M65], osteocalcin, and CTX		X				X		X	
Fibrosure, ELF score, APRI, FIB-4 index, Pro-C3		X				X		X	
Serum sample for NMR analysis		X						X	
Blood sample for genotyping analyses (optional)[11]		X							
Plasma samples for future miRNA and biomarker testing[12]		X				X		X	
Trough PK sample[13]			X		X	X	X	X	

1. ICF may be obtained before the 42-day screening window. Re-testing is allowed within the 42-day Screening period if, in the opinion of the Investigator, the subject may become eligible upon repeat testing.
2. Height and body weight are collected at Screening only; body weight, waist circumference measured midway between the lowest rib and the iliac crest, and signs of peripheral edema are collected at all study visits.
3. Vital sign measurements including seated BP and heart rate are collected at every visit; oral temperature is collected at screening, pre-dose on Day 1 and Visit 8.
4. FibroScan is required within 90 days prior to or during Screening.

5. A qualifying liver biopsy is required within 9 months prior to Screening. If a historical liver biopsy result is not available within this period, a liver biopsy must be performed during Screening for eligible subjects who meet all other entry criteria. A liver biopsy is also performed at Visit 7.
6. Non-diabetic subjects are provided the NCEP dietary guidance at Screening and diabetic subjects are provided the ADA dietary guidance at Screening. Guidance regarding alcohol consumption is also provided to all subjects at Screening. All subjects are assessed for compliance with dietary guidelines and alcohol consumption at each visit.
7. Study staff will contact the subjects by phone every month between Visits 4-7 to inquire about adverse events and serious adverse events.
8. 3 bottles dispensed at Visits 1, 4, 5, and 6.
9. Minimum 10-hour fasting blood sample collected pre-dose.
10. FPG and HbA1c will also be measured at Visits 1, 5 and 7.
11. A separate genotyping informed consent is required to enable collection of blood sample for genotyping analyses.
12. At each timepoint, plasma samples will be split into two separate vials and stored for biomarker testing, including miRNA, to be determined at a later date.
13. Pre-dose trough PK sample collected for analysis of MSDC-0602 and MSDC-0597.