MEDPACE

STATISTICAL ANALYSIS PLAN

Protocol Title:	A PHASE 2A, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED STUDY EVALUATING THE SAFETY AND EFFICACY OF AKR-001 IN SUBJECTS WITH NONALCOHOLIC STEATOHEPATITIS (NASH)
Protocol Number:	AK-US-001-0101
Protocol Version/Date:	Version 1.0 / 26 March 2019 Version 2.0 / 10 April 2019 Amendment 1 / 05 June 2019 Amendment 2 / 13 September 2019 Amendment 3 / 02 October 2019 Amendment 4 / 08 January 2020
Investigational Product:	AKR-001
Sponsor:	Akero Therapeutics, Inc [redacted]
SAP Version/Date:	Version 2.0 / 04 March 2020

CONFIDENTIAL

SIGNATURE PAGE

Protocol Title:	A PHASE 2A, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED STUDY EVALUATING THE SAFETY AND EFFICACY OF AKR-001 IN SUBJECTS WITH NONALCOHOLIC STEATOHEPATITIS (NASH)
Protocol Number:	AK-US-001-0101
SAP Version/Date:	Version 2.0 / 04 March 2020

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

Signature **Date**

[redacted]

VERSION HISTORY

Version	Version Date	Description
1.0	24 January 2020	Original signed version
2.0	04 March 2020	Additional baseline lab parameters to 3.3.4; MELD score derivation description to 3.4.3.1; An additional histology exploratory endpoint to 2.3.3 and 3.4.3.5: ≥ 2 point decrease in NAS score with 0-1 point inflammation, 0 point ballooning, without worsening of fibrosis; An additional exploratory endpoint to 3.4.3.7: non-HDL cholesterol; and Additional exploratory endpoint to 2.3.3 and 3.4.3.9 for MRI-PDFF responders.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation	Definition
ADA	Anti-drug Antibodies
AE	Adverse Event
AESI	Adverse Event of Special Interest
ALP	Alkaline Phosphatase
ALT	Alanine Aminotransferase
ANCOVA	Analysis of Covariance
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Chemical
BLQ	Below Limit of Quantification
BMI	Body Mass Index
BSAP	Bone Specific Alkaline Phosphatase
CLDQ-NAFLD	Chronic Liver Disease Questionnaire-Nonalcoholic Fatty Liver Disease
CP	Child Pugh
CRN	Clinical Research Network
CRF	Case Report Form
CSR	Clinical Study Report
cT1	Corrected T1
CTCAE	Common Terminology Criteria for Adverse Events
DILI	Drug-Induced Liver Injury
DMC	Data Monitoring Committee
DXA	Dual Energy X-ray Absorptiometry
eDISH	Evaluation of Drug-Induced Serious Hepatotoxicity
ECG	Electrocardiogram
ELF™	Enhanced Liver Fibrosis Panel
EOT	End of Treatment
FAS	Full Analysis Set
GGT	Gamma Glutamyl Transferase
HRQoL	Health-Related Quality of Life
INR	International Normalized Ratio
LLOQ	Low Limit of Quantification
MedDRA	Medical Dictionary for Regulatory Activities
MELD	Model for End-Stage Liver Disease
MMRM	Mixed-Model Repeated-Measures
MRI-PDFF	Magnetic Resonance Imaging - Proton Density Fat Fraction
NAB	Neutralizing Antibody
NAFLD	Nonalcoholic Fatty Liver Disease
NAS	NAFLD Activity Score
NASH	Nonalcoholic Steatohepatitis
PK	Pharmacokinetics
PPAS	Per Protocol Analysis Set
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SF-36	Short Form 36 Health Survey
SOC	System Organ Class
TB	Total Bilirubin

Abbreviation	Definition
TEAE	Treatment-Emergent Adverse Event
TESAE	Treatment-Emergent Serious Adverse Event
TFLs	Tables, Figures, and Listings
WHO	World Health Organization
WPAI	Work Productivity and Activity Impairment

1 INTRODUCTION

The purpose of this Statistical Analysis Plan (SAP) is to provide a description of the statistical methods to be implemented for the analysis of the Main Study data from Akero Therapeutics, Inc. Protocol AK-US-001-0101. A separate SAP will be provided for the analysis of Cohort C data from Akero Therapeutics, Inc. Protocol AK-US-001-0101. The SAP will be finalized prior to unblinding of the database. Any deviations from the SAP after database unblinding will be documented in the final Clinical Study Report (CSR).

2 STUDY OVERVIEW

2.1 Study Objectives

2.1.1 Primary Objective

The primary objective of this study is:

- To evaluate absolute change from baseline in hepatic fat fraction assessed by Magnetic Resonance Imaging - Proton Density Fat Fraction (MRI-PDFF) at Week 12

2.1.2 Secondary Objectives

The secondary objectives of this study are:

- To evaluate percent change from baseline in hepatic fat fraction assessed by MRI-PDFF at Week 12
- To evaluate the responder: subjects who achieved a clinically meaningful relative reduction of at least 30% in liver fat content as measured by MRI-PDFF at Week 12
- To evaluate the responder based on Nonalcoholic Fatty Liver Disease (NAFLD) Activity Score system (NAS): subjects who had a decrease of ≥ 2 points in NAS with at least a 1- point reduction in either lobular inflammation or hepatocellular ballooning and with no concurrent worsening of fibrosis stage
- To evaluate the change from baseline in alanine aminotransferase (ALT) at Week 12
- To assess the safety and tolerability of AKR-001 in subjects with NASH

2.1.3 Exploratory Objectives

The exploratory objectives of this study are:

- Change from baseline in markers of liver injury and function including aspartate aminotransferase (AST), alkaline phosphatase (ALP), gamma glutamyl transferase (GGT), bilirubin, International Normalized Ratio (INR), albumin, Model for End-Stage Liver Disease (MELD) score, and Child Pugh (CP) score
- Change from baseline in body weight and waist-to-hip ratio
- MRI-PDFF based quantitation of hepatic fat
- Changes from baseline in noninvasive plasma (or serum) biomarkers including liver fibrosis by Enhanced Liver Fibrosis (ELF™) Test score, and by pro-C3 levels

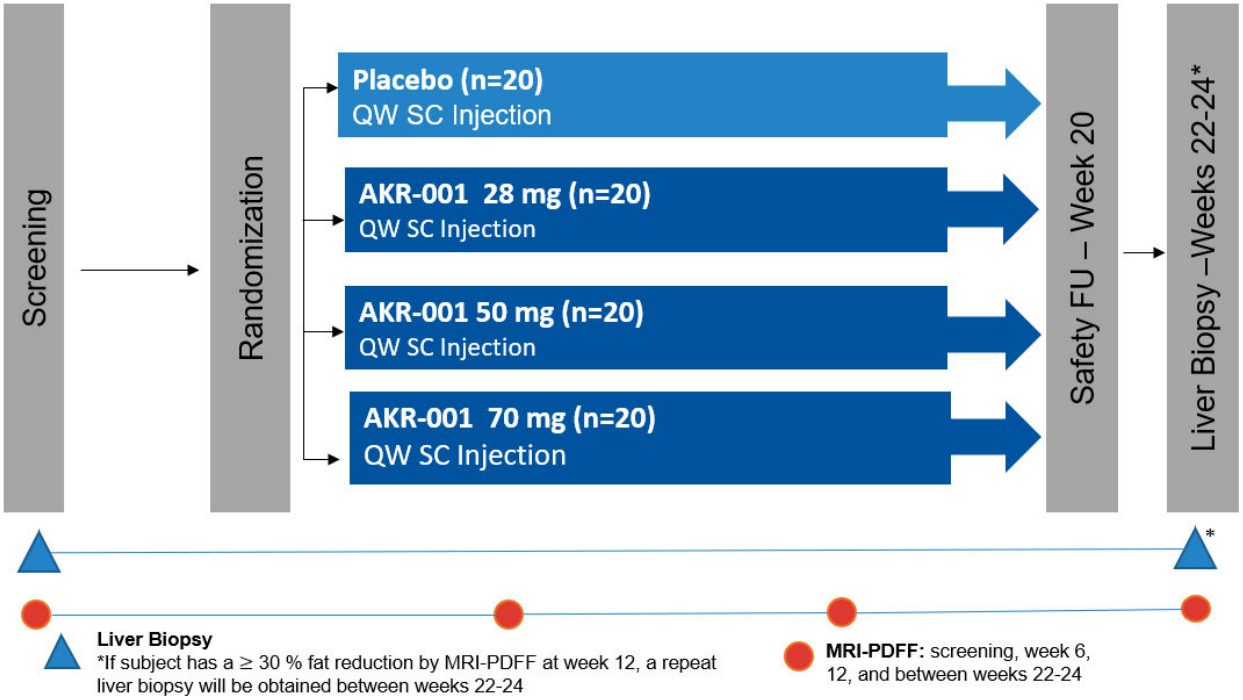
- Change from baseline in bone mineral density of the lumbar spine, femoral neck, and total hip as measured by Dual Energy X-ray Absorptiometry (DXA)
- The proportion of subjects at Week 22 - 24 whose fibrosis regresses by ≥ 1-stage according to the NASH Clinical Research Network (CRN) classification
- The proportion of subjects who had no worsening in NAS with improvement in fibrosis stage
- The proportion of subjects who had 2 points improvement in NAS with no worsening in fibrosis stage
- The proportion of subjects who had ≥ 1 point improvement in NAS, fibrosis, steatosis lobular inflammation or hepatocellular ballooning
- Changes from baseline in health-related quality of life (HRQoL)
- Changes from baseline in lipoprotein profile and markers of lipid metabolism, insulin sensitivity and glycemic control
- Change from baseline in body fat distribution as measured by DXA total body composition
- Change from baseline in corrected T1 (cT1), an MRI-based metric of liver fibro-inflammation at Week 12
- Plasma concentrations of AKR-001

2.2 Study Design

2.2.1 Overview

This is a Phase 2a, randomized, double-blind, placebo-controlled study evaluating the safety and efficacy of AKR-001 in subjects with NASH.

This study will enroll approximately 80 subjects with biopsy proven F1 - F3 NASH. F1 fibrosis will comprise approximately 35% of total subject population. Subjects meeting the study's entry criteria will be randomly assigned in 1:1:1:1 ratio into 4 treatment groups as shown in the figure below:



Randomization will be stratified by baseline hepatic fat fraction (< 15% vs ≥ 15%) and F1 fibrosis score. Study drug will be administered subcutaneously once weekly (QW) for a total of 16 weeks. After 16 weeks of treatment, a safety follow-up visit will occur at Week 20. Subjects with ≥ 30% relative fat reduction on MRI-PDFF at Week 12 will be required to return between Weeks 22 - 24 for a repeat liver biopsy and MRI-PDFF.

This study will be conducted at approximately 30 sites in the U.S. Additional sites may be added, depending on subject accrual rates.

2.2.2 Sample Size Determination

Due to the exploratory nature of this study, no formal power calculations were used to determine sample size. The number of subjects was chosen based on clinical experience with other similar proof of concept studies.

With the following assumptions, the power for N = 20 subjects with active treatment is approximately 91% to detect:

- A true, baseline-adjusted, mean decrease from baseline in MRI-PDFF-estimated hepatic fat of 5% compared to placebo,
- With a one-sided t-test for decrease from placebo,
- A 5% significance level,
- 20 subjects receiving placebo, and
- A standard deviation estimated as 5.2% (Sanyal et al. 2019).

2.3 Study Endpoints

2.3.1 Primary Efficacy Endpoints

The primary endpoint is the absolute change from baseline in hepatic fat fraction measured by MRI-PDFF at Week 12.

2.3.2 Secondary Efficacy Endpoints

The secondary endpoints are:

- Absolute change from baseline in hepatic fat fraction measured by MRI-PDFF at Week 22 - 24
- Percent change from baseline in hepatic fat fraction measured by MRI-PDFF at Weeks 12 and 22 - 24
- Responder: subjects who achieved a clinically meaningful relative reduction of at least 30% in liver fat content as measured by MRI-PDFF at Week 12
- Responder based on NAS system: subjects who had a decrease of ≥ 2 points in NAS with at least a 1-point reduction in either lobular inflammation or hepatocellular ballooning and with no concurrent worsening of fibrosis stage
- Change from baseline in ALT at Weeks 12, 16 and 20

2.3.3 Exploratory Efficacy Endpoints

The exploratory endpoints are:

- Change from baseline in markers of liver injury and function including AST, ALP, GGT, bilirubin, INR, albumin, MELD score, and CP score at Weeks 12, 16 and 20
- Change from baseline in body weight and waist-to-hip ratio at Weeks 12, 16 and 20
- MRI-PDFF based quantitation of hepatic fat
- Changes from baseline in noninvasive biomarkers including Pro-C3 at Weeks 12, 16 and 20, and liver fibrosis by ELF™ Test score at Weeks 6 and 12
- Change from baseline in bone mineral density of the lumbar spine, femoral neck, and total hip as measured by DXA at Week 16
- The proportion of subjects at Week 22 - 24 whose fibrosis regresses by ≥ 1-stage according to the NASH CRN classification
- The proportion of subjects at Week 22 - 24 who had no worsening in NAS with improvement in fibrosis stage
- The proportion of subjects at Week 22 - 24 who had 2 points improvement in NAS with no worsening in fibrosis stage
- The proportion of subjects at Week 22 - 24 who had ≥ 1 point improvement in NAS, fibrosis, steatosis lobular inflammation or hepatocellular ballooning
- The proportion of subjects at Week 22 - 24 who had ≥ 2 point decrease in NAS score with 0-1 point inflammation, 0 point ballooning, without worsening of fibrosis

- Changes from baseline in HRQoL (SF-36, CLDQ-NAFLD, WPAI) at Weeks 12 and 16
- Changes from baseline in lipoprotein profile and markers of lipid metabolism, insulin sensitivity and glycemic control at Weeks 12, 16 and 20
- Change from baseline in body fat distribution as measured by DXA total body composition at Week 16
- Change from baseline in cT1, an MRI-based metric of liver fibro-inflammation at Weeks 12 and 22 - 24
- Plasma concentrations of AKR-001
- The proportion of subjects who achieved a relative reduction of at least 50% in liver fat content as measured by MRI-PDFF at Week 12
- The proportion of subjects who achieved a relative reduction of at least 70% in liver fat content as measured by MRI-PDFF at Week 12
- The proportion of normalized subjects whose liver fat content as measured by MRI-PDFF ≤ 5% at Week 12

2.3.4 Safety Endpoints

The safety endpoint is safety and tolerability of AKR-001 in subjects with NASH.

Safety analyses include: summaries of extent of exposure, adverse events (AEs), laboratory evaluations, electrocardiogram (ECG), vital sign assessments, body weight, anti-drug antibody (ADA) and neutralizing antibody (NAB) assessments, and concomitant medication usage.

3 STATISTICAL METHODOLOGY

3.1 General Considerations

3.1.1 Analysis Day

Analysis day will be calculated from the date of first dose of study drug. The day of the first dose of study drug will be Day 1, and the day immediately before Day 1 will be Day -1. There will be no Day 0, in alignment with international data standards for clinical trials. Analysis day will be calculated as (date of clinical event – date of first dose of study drug) + 1 if the date of the event is on or after the date of first dose and as (date of clinical event – date of first dose of study drug) if the date of event is before the date of first dose.

3.1.2 Analysis Visits

Scheduled visits will be assigned to analysis visits as recorded on the Case Report Form (CRF). Unscheduled and early termination visits for the MRI-PDFF, cT1, ELF™ and pro-C3 will be assigned to analysis visits according to the following visit windows:

Analysis Visit	Target Analysis Day	Low Analysis Day	High Analysis Day
Week 6	43	22	64
Week 12	85	65	127
Note: if the unscheduled collection or early term collection is >21 days after the last dose of study drug, it will not be used in analyses.			

For ALT, markers of liver injury and function, waist-to-hip ratio, biomarkers and HRQoL, unscheduled and early termination visits will be assigned to analysis visits according to the following visit windows:

Analysis Visit	Target Analysis Day	Low Analysis Day	High Analysis Day
Week 1	8	2	15
Week 4	29	16	43
Week 8	57	44	71
Week 12	85	72	99
Week 16	113	100	127
Week 20 Safety follow-up	141	128	EOS
Note: if the unscheduled collection or early term collection is >14 days after the last dose of study drug, it will not be used in analyses.			

For body weight, unscheduled and early termination visits will be assigned to analysis visits according to the following visit windows:

Analysis Visit	Target Analysis Day	Low Analysis Day	High Analysis Day
Week 1	8	2	11
Week 2	15	12	18
Week 3	22	19	25
Week 4	29	26	32
Week 5	36	33	39
Week 6	43	40	46
Week 7	50	47	53
Week 8	57	54	60
Week 9	64	61	67
Week 10	71	68	74
Week 11	78	75	81
Week 12	85	82	88
Week 13	92	89	95
Week 14	99	96	102
Week 15	106	103	109
Week 16	113	110	127
Week 20 Safety follow-up	141	128	EOS

The efficacy assessments from two biopsies (Screening and Repeat Biopsy Weeks 22 - 24) and the safety assessments from two DXA Scans (Screening and Week 16) will be considered for the analysis regardless of when the second assessment or biopsy takes place. Unscheduled visits after the first dose of study drug and early termination visits will be assigned as post-baseline.

For PK concentration summaries, a concentration that is intended as predose will be excluded from analysis if it is >24 hours before a dose or after the current dose. Beyond this rule, PK concentrations must be within the visit window to be included in summary statistics of PK.

Within an analysis visit window, the measurement from the scheduled visits will be used if available. If no scheduled visit occurs, the measurement from the visit closest to the defined target analysis day within the window will be used. If there is more than one measurement with

equal distance to the defined target analysis day, the latter will be used. If no visits occur within a visit window, the measurement for this visit will be treated as missing.

3.1.3 Definition of Baseline

For all variables except lipids and liver enzymes variables, pre-dose measurements on Baseline visit (Day 1) will be considered the baseline values. If the measurement at this visit is not available, the last measurement prior to the first dose of study drug may be used as the baseline value unless it is stated otherwise. The scheduled baseline visit for MRI-PDFF, FibroScan, liver biopsy, DXA scan, cT1, ELF™ and pro-C3 is Screening. Baseline lipid variables including Total Cholesterol, HDL-C, LDL-C, Triglycerides, non-HDL cholesterol will be the average of Screening and Baseline visit (Day 1) values. Baseline liver enzymes variables including ALT and AST will be the average of Pre-Baseline and Baseline visit (Day 1) values. If one is missing for lipids and liver enzymes variables, the available one will be used.

3.1.4 Summary Statistics

Categorical data will generally be summarized with counts and percentages of subjects. The denominator used for the percentage calculation will be clearly defined. Continuous data will generally be summarized with descriptive statistics including n (number of non-missing values), mean, median, standard deviation, minimum, and maximum. For certain continuous data, it will be summarized with descriptive statistics including n (number of non-missing values), arithmetic mean, arithmetic standard deviation, geometric mean, geometric coefficient of variation, median, minimum, and maximum. PK concentrations will be summarized with the set of continuous parameters including geometric mean and will add the statistic of number of samples below the limit of quantification (nblq).

All individual subject data will be listed as measured.

3.1.5 Hypothesis Testing

The statistical hypotheses will be tested at the 2-sided, 0.05 significance for the parameters with neutral direction other than specified. Fixed hypothesis testing sequence will be used to control the multiplicity for each active dose to be compared with placebo.

3.1.6 Handling of Dropouts and Missing Data

Missing data can have an impact on the interpretation of the trial data. In general, values for missing data will not be imputed.

Where appropriate, safety data for subjects that did not complete the study will be included in summary statistics. For example, if a subject received study drug, the subject will be included in a summary of AEs according to the treatment received; otherwise, if the subject is not dosed then they will be excluded from the summary. If safety laboratory results for a subject are missing for any reason at a time point, the subject will be excluded from the calculation of summary statistics for that time point. If the subject is missing a pre-dose value, then the subject will be excluded from the calculation of summary statistics for the pre-dose value and the change from pre-dose values.

Values for missing safety laboratory data will not be imputed; however, a missing baseline result will be replaced with a screening result, if available. If no pre-treatment laboratory value is available, the baseline value will be assumed to be normal (i.e., no grade [Grade 0]) for the summary of graded laboratory abnormalities.

Values for missing vital signs data will not be imputed; however, a missing baseline result will be replaced with a screening result, if available.

Below limit of quantification (BLQ) PK concentration at pre-dose at Baseline will be imputed as zero and the other pre-dose BLQ PK concentrations will be imputed as half of the low limit of quantification (0.5*LLOQ).

3.2 Analysis Populations

3.2.1 Full Analysis Set (FAS)

The FAS will include all subjects who were randomized into the study. The FAS will be the primary analysis set for efficacy analyses.

Subjects who receive a dose of study drug other than that to which they were randomized for the entire duration of treatment will be analyzed according to treatment group to which they were randomized.

3.2.2 Safety Set

The Safety Set will include all subjects who received at least one dose of study drug. The Safety Set will be the primary analysis set for safety analyses.

Subjects who received a dose of study drug other than that to which they were randomized for the entire duration of treatment will be analyzed according to the dose of study drug received. All data collected during treatment plus 30 days after last dose of study drug will be included in the safety summaries.

3.2.3 PK Analysis Set

The PK Analysis Set will include all randomized subjects who took at least one dose of study drug and for whom concentration data of AKR-001 is available.

3.2.4 Biomarker Analysis Set

The Biomarker Analysis Set will include data from subjects in the Safety Set who have the necessary baseline and on-study measurements to provide interpretable results for the specific parameters of interest.

3.2.5 Per Protocol Analysis Set (PPAS)

The PPAS is defined as any subject who is randomized, receives at least one dose of study drug, and has Baseline and Week 12 hepatic fat fraction assessed by MRI-PDFF without any major protocol deviations. A full detail of inclusions and exclusions from PPAS will be in the classifications specifications and finalized before unblinding.

3.2.6 MRI-PDFF Evaluable Analysis Set

MRI-PDFF Evaluable Analysis Set will be a subset of the FAS. It will include all FAS subjects who have Baseline and Week 12 hepatic fat fraction assessed by MRI-PDFF. It will be the analysis population for the responders based on MRI-PDFF.

3.2.7 Liver Biopsy Evaluable Analysis Set

Liver Biopsy Evaluable Analysis Set will be a subset of the FAS. It will include all FAS subjects who have Baseline and Week 22 - 24 liver biopsy results. It will be the analysis population for the responders based on NAS system and fibrosis stage.

3.2.8 Stratification

For statistical analysis, the factors on which randomization has been stratified will be used. Preservation of the initial randomization in analysis is important in preventing bias and in providing a secure foundation for statistical tests (ICH E9).

3.3 Subject Data and Study Conduct

3.3.1 Subject Disposition

Counts and percentages of subjects who were screened (signed informed consent), discontinued early during screening (screen failures), and randomized will be summarized in total based on all screened subjects. Reasons for early discontinuation will also be summarized.

Counts and percentages of subjects in each of the following disposition categories will be presented by treatment and in total based on all randomized subjects:

- Subjects who were randomized
- Subjects who started study drug
- Subjects who completed the 12-week treatment
- Subjects who completed the 16-week treatment
- Subjects who withdrew from the 16-week treatment
- Subjects who completed the study
- Subjects who withdrew from the study

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

3.3.2 Protocol Deviations

Protocol deviations will be defined in the Protocol Deviation Plan. Counts and percentages of subjects with CSR reportable protocol deviations by deviation category will be summarized by treatment and in total based on all randomized subjects.

3.3.3 Analysis Populations

Counts and percentages of subjects in each analysis population will be summarized by treatment and in total based on all randomized subjects.

3.3.4 Demographic and Baseline Characteristics

The following demographic and baseline characteristics will be summarized:

- Age (years) and age categories (<45 years, ≥45 to <65 years, ≥65 years)
- Sex

- Race
- Ethnicity
- Height (cm) at Screening
- Weight (kg) at Baseline
- Body mass index (BMI) (kg/m^2) at Baseline and BMI categories at Baseline (<30 kg/m^2, ≥30 kg/m^2)
- Waist circumference (cm) at Baseline
- Hip circumference (cm) at Baseline
- Waist-to-hip ratio at Baseline
- Type 2 diabetes status at Screening
- Fibrosis stage at Screening
- Biopsy diagnosis at Screening
- NAS and NAS components at Screening
- Hepatic fat fraction (%) by MRI-PDFF at Screening
- Stratification groups at Baseline: F1 fibrosis score (F1, F2-3) and hepatic fat fraction categories (< 15%, ≥ 15%)
- Alanine aminotransferase (U/L) and Aspartate Aminotransferase (U/L) at Baseline
- Triglyceride (mg/dL), Total Cholesterol (mg/dL), LDL-C (mg/dL), HDL-C (mg/dL), and non-HDL-C (mg/dL) at Baseline
- Hemoglobin A1c (%) at Baseline
- Endogenous FGF21 at Baseline
- Bone Mineral Density at Baseline
- Body Fat Distribution at Baseline

Demographic and baseline characteristics will be summarized with descriptive statistics or counts and percentages of subjects as appropriate by treatment and in total for all randomized subjects and each defined analysis population if the analysis population composition is different than all randomized subjects.

3.3.5 Medical History

Medical history will be coded to system organ class (SOC) and preferred term using the Medical Dictionary for Regulatory Activities (MedDRA) version 22.0 or the latest version. Counts and percentages of subjects with medical history by SOC and preferred term will be summarized by treatment and in total based on all randomized subjects.

3.3.6 Concomitant Medications

Concomitant medications will be coded to anatomical therapeutic chemical (ATC) class and preferred term using the World Health Organization (WHO) Drug Dictionary version 2019G B3 or the latest version. For summary purposes, medications will be considered prior medications if they stopped prior to the first dose of study drug and concomitant medications if they were taken at any time on or after the first dose of study drug (i.e. started prior to the first dose of study drug and were ongoing or started after the first dose of study drug).

Counts and percentages of subjects taking concomitant medications by ATC class and preferred term will be summarized by treatment and in total based on the Safety Set. The use of any prior medication or concomitant medication will be listed for all randomized subjects.

3.3.7 Study Drug Exposure and Compliance

Weeks of exposure to study drug will be calculated as:

(date of last dose of study drug - date of first dose of study drug + 1) / 7

Weeks of exposure to study drug will be summarized by treatment based on the Safety Set with descriptive statistics and with counts and percentages of subjects with exposure in the following categories:

- >0 to ≤4 weeks
- >4 to ≤8 weeks
- >8 to ≤12 weeks
- >12 to ≤16 weeks
- >16 weeks

The total number of injections administered will be calculated for each subject. A contingency table will be provided to display the number and percentage of subjects in each treatment group with the following number of injections: 1-3, 4-6, 7-9, 10-12, 13-15, >=16.

Subjects that miss 2 or more doses of study drug over the course of study participation are considered noncompliant. The frequency and percentage of subjects who are compliant and noncompliant will be summarized by treatment group.

3.4 Efficacy Assessment

Efficacy data will be summarized and analyzed by randomized treatment based on the FAS unless otherwise specified.

3.4.1 Primary Efficacy Endpoint

The primary efficacy endpoint will be the absolute change from baseline hepatic fat fraction measured by MRI-PDFF at Week 12.

Absolute change in fat fraction by MRI-PDFF = Week 12 - Baseline

MRI-PDFF will be performed at Screening, Week 6 and Week 12, and a repeat MRI-PDFF in eligible subjects will be performed at Week 22 - 24. Summary statistics at all visits for the values, change from baseline and percent change from baseline will be provided.

3.4.1.1 Primary Analysis

The primary efficacy analysis will be analyzed with an analysis of covariance (ANCOVA) model with treatment group and F1 fibrosis score (F1 vs F2-3) as factors and baseline hepatic fat fraction measured by MRI-PDFF as a covariate. The SAS sample code is listed:

```
*****************************************************************
  TREATMENT: 0 (Placebo), 1 (AKR-001 28 mg), 2 (AKR-001 50 mg), or
             3 (AKR-001 70 mg)
  FSCORE: Fibrosis score (F1 vs F2-3)
  BASE: Baseline fat fraction by MRI-PDFF
  CHG: Change from baseline to Week 12 in fat fraction by MRI-PDFF
*****************************************************************;
proc glm;
```

```
  class TREATMENT FSCORE;
  model CHG = TREATMENT FSCORE BASE / p;
  lsmeans TREATMENT / stderr pdiff cl;
  estimate "AKR-001 70 mg : Placebo" TREATMENT -1 0 0 1;
  estimate "AKR-001 50 mg : Placebo" TREATMENT -1 0 1 0;
  estimate "AKR-001 28 mg : Placebo" TREATMENT -1 1 0 0;
run;
```

For any subjects in the FAS with a missing primary efficacy value, the primary efficacy variable will be imputed with multiple imputation method.

The treatment effect of each active dose comparing with placebo will be estimated from the ANCOVA model. LS means, one-sided, 97.5% confidence intervals and one-sided p-values will be calculated. Normality will be tested for the model residuals by the Shapiro-Wilk normality test.

In order to control the Type I error rate inflation caused by multiplicity issues, a step-down fixed hypothesis testing sequence will be used, following the order of AKR-001 70 mg vs. Placebo, AKR-001 50 mg vs. Placebo, and then AKR-001 28 mg vs. Placebo. That is, only if the higher dose is proved to be significantly better than placebo will the next low dose be compared.

Multiple Imputation Method:

Missing primary efficacy endpoint will be handled using the MAR multiple imputation method.

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

- Multiple imputation will be used to impute missing data and will be implemented using SAS® PROC MI: Missing data may be the result of missing hepatic fat fraction (%) assessed by MRI-PDFF at Week 12 or the result of subjects discontinuing treatment prior to Week 12. Missing at random (MAR) will be assumed. The imputation model will include the subject demographics (age [<median, ≥median], race [white, non-white], baseline weight [<median, ≥median]), F1 fibrosis score [F1, F2-3], and baseline and Week 6 hepatic fat fraction assessed by MRI-PDFF. Missing data will be imputed 100 times to generate 100 complete data sets. The seed will be 98765. The fully conditional specification method will be used. The 100 complete data sets will then be analyzed using the same analysis method as the one used to analyze the primary endpoint.
- The results from the 100 fitted models will be combined using SAS® PROC MIANALYZE: The estimate of the treatment difference and standard error from the analysis will be analyzed by PROC MIANALYZE to obtain the overall estimate of treatment difference, as well as the confidence interval and p-value for the hypothesis testing.
- If the model fails to converge, exclusion of subject demographics will be considered.

3.4.1.2 Secondary Analysis

The absolute change from baseline in hepatic fat fraction measured by MRI-PDFF will be analyzed using a mixed-model repeated-measures (MMRM) on the FAS. No imputation will be performed. The factors in the model will be F1 fibrosis score, treatment group, baseline value, visit, and the treatment group by visit interactions. An unstructured covariance matrix will be used. The SAS sample code is listed:

```
*************************************************************************
  USUBJID: Unique subject identifier
  TREATMENT: 0 (Placebo), 1 (AKR-001 28 mg), 2 (AKR-001 50 mg), or
             3 (AKR-001 70 mg)
  FSCORE: Fibrosis score (F1 vs F2-3)
  BASE: Baseline fat fraction by MRI-PDFF
  VISIT: Week 6, Week 12
  CHG: Change from baseline to Week 12 in fat fraction by MRI-PDFF
*************************************************************************;
proc mixed;
  class USUBJID FSCORE TREATMENT VISIT;
  model CHG = FSCORE TREATMENT BASE VISIT TREATMENT*VISIT;
  Repeated VISIT / TYPE=UN sub=USUBJID;
  lsmeans VISIT*TREATMENT / cl pdiff slice=VISIT;
run;
```

3.4.1.3 Sensitivity Analyses

- Robustness of the results to departure from the MAR assumption will be explored in the FAS using tipping-point analysis approach: this sensitivity analysis assumes the "missing not at random (MNAR)" mechanism for dropouts in the active treatment group and the "missing at random (MAR)" mechanism for dropouts in the placebo group in this study. It also assumes that dropouts from the active treatment group at a given post-baseline visit would have, on average, their unobserved fat fraction worse by δ compared to the fat fraction imputed by a Multiple Imputation (MI) approach based on the MAR method.
  - Estimations will be performed using the multiple imputation approach as described above. In each treatment group, a penalty δ will be added to the imputed values (δ = 0 corresponds to the MAR assumption) in the multiple imputation step.
  - To investigate how the conclusions for the endpoint depend on the adopted values δ, the testing will be repeated over a range of plausible values for each treatment groups. Results will then be summarized using contour plot.

    The smallest value of $\delta_{\_treatment}$ while $\delta_{\_placebo} = 0$ which will statistically reverse the conclusion for the primary endpoint will be assessed and interpreted from its clinical plausibility.
- The primary analysis will be repeated on the PPAS. No imputation will be performed.
- The treatment group-by-baseline MRI-PDFF interaction will be added in the ANCOVA model of the primary analysis to check the homogeneity of regression slopes assumption.

3.4.2 Secondary Efficacy Endpoints

3.4.2.1 Absolute change from baseline in fat fraction measured by MRI-PDFF at Week 22 - 24

This analysis is defined for subjects with ≥ 30% relative fat reduction on MRI-PDFF at Week 12 and required to return between Weeks 22 - 24. ANCOVA model with treatment group, F1 fibrosis score (F1 vs F2-3) as factors and baseline hepatic fat fraction as a covariate will be performed on the FAS. LS means, 95% confidence intervals and p-values will be presented. The same MI method will be used for the missing value at Week 22 - 24. The analysis will be

repeated on the observed data only including subjects with non-missing baseline and Weeks 22 - 24 fat fraction.

3.4.2.2 Percent change from baseline in fat fraction measured by MRI-PDFF at Weeks 12 and 22 - 24

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

For the percent change from baseline to Week 12, the same ANCOVA model used for primary efficacy endpoint will be performed on the FAS with the same multiple imputation method, and repeated on the PPAS without imputation, as well as the same MMRM model on the FAS. The same statistics will be presented as for the primary model. Normality will be tested for the model residuals by the Shapiro-Wilk normality test. The homogeneity of regression slopes assumption will be checked. For the percent change from baseline to Week 22 - 24, the analysis is defined for subjects with ≥ 30% relative fat reduction on MRI-PDFF at Week 12 and required to return between Weeks 22 - 24. The same MI method will be used for the missing value at Week 22 - 24. The analysis will be repeated on the observed data only including subjects with non-missing baseline and Weeks 22 - 24 fat fraction.

3.4.2.3 Responder based on MRI-PDFF: subjects who achieved a clinically meaningful relative reduction of at least 30% in liver fat content as measured by MRI-PDFF at Week 12

That is the percent change from baseline to Week 12 ≤ -30% in hepatic fat fraction. The number and percentage of responders will be provided by treatment group and by stratification group. Logistic regression will be used for the analyses based on FAS with missing values imputed as non-responders and repeated on MRI-PDFF Evaluable Analysis Set without imputation. The analyses will include the treatment group and F1 fibrosis score (F1 vs F2-3) as factors and baseline hepatic fat fraction measured by MRI-PDFF as a covariate. The odds ratio, 95% confidence interval, and p-value from the logistic regression will be provided. The SAS sample code is listed:

```
*********************************************************************
  TREATMENT: 0 (Placebo), 1 (AKR-001 28 mg), 2 (AKR-001 50 mg), or
             3 (AKR-001 70 mg)
  FSCORE: Fibrosis score (F1 vs F2-3)
  BASE: Baseline fat fraction by MRI-PDFF
  RESPONSE: 0 (Non-responder), 1 (Responder)
*********************************************************************;
proc logistic;
  class TREATMENT FSCORE;
  model RESPONSE (event='1') = TREATMENT FSCORE BASE / link=logit
        clodds=wald orpvalue;
  oddsratio TREATMENT /diff=ref;
run;
```

3.4.2.4 Change from baseline in ALT at Weeks 12, 16 and 20

ALT will be measured at Screening, Pre-Baseline, Baseline (Day 1), Week 1, Week 4, Week 8, Week 12, Week 16 and Week 20 Safety follow-up. Values and changes from baseline in ALT will be summarized descriptively. ANCOVA model with treatment group, baseline hepatic fat fraction (< 15% vs ≥ 15%), and F1 fibrosis score (F1 vs F2-3) as factors and baseline value as a covariate will be performed on the FAS with the same multiple imputation method in Section 3.4.1.1, and repeated on the PPAS without imputation, as well as the same MMRM model on the FAS in Section 3.4.1.2. LS means, 95% confidence intervals and p-values will be

presented. Normality will be tested for the model residuals by the Shapiro-Wilk normality test. The homogeneity of regression slopes assumption will be checked.

3.4.2.5 Responder based on NAS system: subjects who had a decrease of ≥ 2 points in NAS with at least a 1-point reduction in either lobular inflammation or hepatocellular ballooning and with no concurrent worsening of fibrosis stage

The responders are defined for subjects with ≥ 30% relative fat reduction on MRI-PDFF at Week 12 and required to return between Weeks 22 - 24. The number and percentage of responders will be provided by treatment group. Fisher's exact test will be used for the analysis using FAS with missing values imputed as non-responders and repeated on Liver Biopsy Evaluable Analysis Set without imputation. Odds ratios (vs Placebo), 95% confidence intervals and p-values will be provided.

3.4.3 Exploratory Efficacy Endpoints

The exploratory efficacy endpoints are defined in Section 2.3.3. Descriptive statistics at each scheduled visit on the values and changes from baseline of the exploratory efficacy endpoints will be summarized. For certain efficacy endpoints, logarithmic transformation may be performed prior to fitting the ANCOVA model. Since these endpoints are for exploratory purposes, no adjustments for multiplicity will be made.

3.4.3.1 Markers of liver injury and function

Markers of liver injury and function including AST, ALP, GGT, bilirubin, INR, albumin, MELD score, and CP score will be measured at Screening, Pre-Baseline, Baseline (Day 1), Week 1, Week 4, Week 8, Week 12, Week 16 and Week 20 Safety follow-up.

MELD score will be calculated as below:

- Candidates who are at least 12 years old receive an initial MELD (i) score equal to: 0.957 x Loge(creatinine mg/dL) + 0. 378 x Loge (bilirubin mg/dL) + 1.120 x Loge (INR) + 0.643
- Multiply the score by 10 and round to the nearest whole number. Laboratory values less than 1.0 are set to 1.0 for the purposes of the MELD score calculation.
- For candidates with an initial MELD score greater than 11, the MELD score is then re-calculated as follows:

  MELD = MELD(i) + 1.32*(137-Na) – [0.033*MELD(i)*(137-Na)], where Sodium values less than 125 mmol/L will be set to 125, and values greater than 137 mmol/L will be set to 137.

CP score will be calculated as below, ranging from 5-15:

	1 point	**2 point**	**3 point**
Bilirubin (Total)	<2 mg/dL (<34.2 µmol/L)	2-3 mg/dL (34.2-51.3 µmol/L)	>3 mg/dL (>51.3 µmol/L)
Albumin	>3.5 g/dL (>35 g/L)	2.8-3.5 g/dL (28-35 g/L)	<2.8 g/dL (<28 g/L)
INR	<1.7	1.7-2.2	>2.2
Ascites	Absent	Slight	Moderate
Encephalopathy	No Encephalopathy	Grade 1-2	Grade 3-4
Class A = 5-6 points Class B = 7-9 points Class C = 10-15 points			

ANCOVA model with treatment group, baseline hepatic fat fraction (< 15% vs ≥ 15%), and F1 fibrosis score (F1 vs F2-3) as factors and baseline value as a covariate will be used for analyzing the change from baseline to Week 12, the change from baseline to Week 16, and the change from baseline to Week 20. LS means, 95% confidence intervals and p-values will be presented. The missing value will be imputed by last-observed-carried-forward (LOCF) method.

3.4.3.2 Body weight and waist-to-hip ratio

Body weight, waist circumference and hip circumference will be measured at all visits.

ANCOVA model with treatment group, baseline hepatic fat fraction (< 15% vs ≥ 15%), and F1 fibrosis score (F1 vs F2-3) as factors and baseline value as a covariate will be used for analyzing the change from baseline to Week 12, the change from baseline to Week 16, and the change from baseline to Week 20. LS means, 95% confidence intervals and two-sided p-values will be presented for body weight, and LS means, 95% confidence intervals and p-values will be presented for the other parameters. The missing value will be imputed by LOCF method.

3.4.3.3 Biomarkers

ELF™ will be performed at Screening, Week 6 and Week 12. Fibrosis biomarker pro-C3, lipid metabolism biomarker and insulin sensitivity & glycemic control biomarker will be measured at Baseline (Day 1), Week 1, Week 4, Week 8, Week 12, Week 16 and Week 20 Safety follow-up.

ANCOVA model with treatment group, baseline hepatic fat fraction (< 15% vs ≥ 15%), and F1 fibrosis score (F1 vs F2-3) as factors and baseline value as a covariate will be used for analyzing the change from baseline to Week 6 and the change from baseline to Week 12 for the ELF™ test score. For other biomarkers, the same ANCOVA model will be used for analyzing the change from baseline to Week 12, the change from baseline to Week 16, and the change from baseline to Week 20. LS means, 95% confidence intervals and p-values will be presented. The missing value will be imputed by LOCF method.

Additional exploratory analyses that could enhance the understanding of the biological effects and the mechanism of action of AKR-001 may be added in the Biomarker Analysis Plan if necessary.

3.4.3.4 DXA scan measurements

DXA scans measurements including bone mineral density of the lumbar spine, femoral neck, total hip and body fat distribution will be performed at Baseline (Day 1) and Week 16.

An analysis will be performed on the change from baseline to Week 16 in DXA Scan measurements using ANCOVA model with treatment group, baseline hepatic fat fraction (< 15% vs ≥ 15%), and F1 fibrosis score (F1 vs F2-3) as factors and baseline value as a covariate. LS means, 95% confidence intervals and two-sided p-values will be presented. The missing value will be imputed by LOCF method.

3.4.3.5 Liver fibrosis responders

- *The proportion of subjects at Week 22 - 24 whose fibrosis regresses by ≥ 1-stage according to the NASH CRN classification;*
- *The proportion of subjects at Week 22 - 24 who had no worsening in NAS with improvement in fibrosis stage;*

- *The proportion of subjects at Week 22 - 24 who had 2 points improvement in NAS with no worsening in fibrosis stage;*
- *The proportion of subjects at Week 22 - 24 who had ≥ 1 point improvement in NAS, fibrosis, steatosis lobular inflammation or hepatocellular ballooning; and*
- *The proportion of subjects at Week 22 - 24 who had ≥ 2 point decrease in NAS score with 0-1 point inflammation, 0 point ballooning, without worsening of fibrosis.*

The responders are defined as subjects with a ≥ 30% relative fat reduction on MRI-PDFF at Week 12 and required to return between Weeks 22 - 24. Liver biopsy will be obtained at Screening and Week 22 - 24. The number and percentage of the responders will be provided by treatment group. Fisher's exact test will be used for the analyses on the Liver Biopsy Evaluable Analysis Set. Odds ratios, 95% confidence intervals and p-values will be provided.

3.4.3.6 HRQoL

HRQoL including Short Form 36 Health Survey (SF-36), Chronic Liver Disease Questionnaire-Nonalcoholic Fatty Liver Disease (CLDQ-NAFLD) and Work Productivity and Activity Impairment (WPAI) will be collected at Baseline (Day 1), Week 1, Week 4, Week 8, Week 12 and Week 16.

ANCOVA model with treatment group, baseline hepatic fat fraction (< 15% vs ≥ 15%), and F1 fibrosis score (F1 vs F2-3) as factors and baseline value as a covariate will be used for analyzing the change from baseline to Week 12 and the change from baseline to Week 16. LS means, 95% confidence intervals and p-values will be presented. The missing value will be imputed by LOCF method.

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

The scores are weighted sums of the questions in each domain. All the raw scores will be transformed to scores ranging from 0 - 100. Higher scores represent better health. Physical Component Summary (PCS) is calculated by positively weighting the four subscales in the physical domain (PF, RP, BP and GH) and the remaining psychological domain subscales negatively. In contrast, the Mental Component Summary (MCS) is calculated by positively weighting the four mental domain subscales (VT, SF, RE and MH), and negatively weighting the four physical domain subscales.

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

CLDQ-NAFLD: The CLDQ-NAFLD asks questions related to liver disease and specifically NAFLD, to measure health related quality of life in subjects with chronic liver disease.

The questionnaire has 36 items grouped into six domains: Abdominal Symptoms, Activity, Emotional, Fatigue, Systemic Symptoms, and Worry. Items are scored by a 7-point rating scale from 1 (all of the time) to 7 (none of the time). The higher item values indicate better health.

The domain score will be calculated as dividing the sum of the domain's items by the number of items within that domain. The total score will be calculated as an average of the six domain scores.

Domains	Number of Items	Items
Abdominal Symptoms (AS)	3	1, 5, 17
Fatigue (FA)	6	2, 4, 8, 11, 13, 35
Systemic Symptoms (SS)	6	3, 6, 21, 23, 27, 36
Activity (AC)	5	7, 9, 14, 30, 31
Emotional Function (EF)	9	10, 12, 15, 16, 19, 20, 24, 26, 34
Worry (WO)	7	18, 22, 25, 28, 29, 32, 33

WPAI: The questionnaire asks questions regarding the effect of NASH on a person's ability to work and perform regular activities.

The WPAI consists of six questions: 1 = currently employed; 2 = hours missed due to specified problem; 3 = hours missed other reasons; 4 = hours actually worked; 5 = degree problem affected productivity while working; 6 = degree problem affected regular activities. WPAI outcomes are expressed as impairment percentages, with higher numbers indicating greater impairment and less productivity. Four main outcomes can be generated and expressed in percentages by multiplying the following scores by 100:

- Percent work time missed due to problem: Q2 / (Q2 + Q4)
- Percent impairment while working due to problem: Q5 / 10
- Percent overall work impairment due to problem:
  Q2 / (Q2 + Q4) + [(1 - (Q2 / (Q2 + Q4))) x (Q5 / 10)]
  For those who missed work and did not actually work in the past seven days, the percent overall work impairment due to problem will be equal to the percent work time missed due to problem.
- Percent activity impairment due to problem: Q6 / 10

3.4.3.7 Lipoprotein profile and free fatty acids

Lipoprotein profile including Total Cholesterol, HDL-C, LDL-C, Triglycerides, and free fatty acids will be measured at Screening, Baseline (Day 1), Week 1, Week 4, Week 8, Week 12, Week 16 and Week 20 Safety follow-up.

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

ANCOVA model with treatment group, baseline hepatic fat fraction (< 15% vs ≥ 15%), and F1 fibrosis score (F1 vs F2-3) as factors and baseline value as a covariate will be used for analyzing the change from baseline to Week 12, the change from baseline to Week 16, and the change from baseline to Week 20. LS means, 95% confidence intervals and p-values will be presented. The missing value will be imputed by LOCF method.

3.4.3.8 cT1

cT1, an MRI-based metric of liver fibro-inflammation, will be performed at Screening, Week 6, Week 12 and Week 22 - 24, where available.

ANCOVA model with treatment group, baseline hepatic fat fraction (< 15% vs ≥ 15%), and F1 fibrosis score (F1 vs F2-3) as factors and baseline value as a covariate will be used for analyzing the change from baseline to Week 12 and the observed change from baseline to Week 22 - 24. LS means, 95% confidence intervals and p-values will be presented. The missing value will be imputed by LOCF method for the change from baseline to Week 12. No

imputation will be performed for the analysis of the observed change from baseline to Week 22 - 24. Only subjects with non-missing baseline and Week 22 - 24 cT1 will be included. The analysis will be performed if data is available.

3.4.3.9 MRI-PDFF responders

- *The proportion of subjects who achieved a relative reduction of at least 50% in liver fat content as measured by MRI-PDFF at Week 12;*
- *The proportion of subjects who achieved a relative reduction of at least 70% in liver fat content as measured by MRI-PDFF at Week 12; and*
- *The proportion of normalized subjects whose liver fat content as measured by MRI-PDFF ≤ 5% at Week 12.*

The number and percentage of responders will be provided by treatment group and by stratification group. Logistic regression will be used for the analyses based on FAS with missing values imputed as non-responders and repeated on MRI-PDFF Evaluable Analysis Set without imputation. The analyses will include the treatment group and F1 fibrosis score (F1 vs F2-3) as factors and baseline hepatic fat fraction measured by MRI-PDFF as a covariate. The odds ratio, 95% confidence interval, and p-value from the logistic regression will be provided.

3.4.4 Subgroups

Subgroup analyses will be performed for the primary efficacy endpoint. The group variables include baseline hepatic fat fraction, F1 fibrosis score, Type 2 diabetes status and baseline Endogenous FGF21.

3.5 Safety Assessment

The safety of AKR-001 will be assessed during the study through the reporting of AEs, clinical laboratory tests, ECG, vital sign assessments, body weight, ADA and NAB assessments, and concomitant medication usage.

All safety data collected on or after the date that study drugs were first dispensed up to the date of last dose of study drugs plus 30 days will be summarized by treatment group. Data for the follow-up period will be included in data listings.

Safety data will be summarized by actual treatment received (and in total for selected analyses) based on the Safety Set unless otherwise specified.

3.5.1 Adverse Events

All AEs will be coded to system organ class and preferred term using MedDRA version 22.0 or the latest version.

Treatment-emergent adverse events (TEAEs) are defined as one or both of the following:

- Any AEs with an onset date/time on or after the study drugs start date/time and no later than 30 days after permanent discontinuation of study drugs
- Any AEs leading to premature discontinuation of study drugs.

Injection site reactions, diarrhea, hypoglycemia and drug-induced liver injuries (DILI) will be monitored as AEs of special interests (AESIs) during this study. During the course of the study, additional AESIs may be identified by the Sponsor.

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

- Any TEAEs (overall and by maximum severity)
- Any study drug related TEAEs (overall and by maximum severity)
- Any study procedure related TEAEs (overall and by maximum severity)
- Any TEAEs of special interest (overall, injection site reactions, diarrhea, hypoglycemia, DILI)
- Any serious AEs (SAEs)
- Any treatment-emergent serious AEs (TESAEs)
- Any TEAEs leading to premature discontinuation of study drug
- Any TEAEs leading to premature discontinuation of study
- Any AEs leading to death

Counts and percentages of subjects will also be presented by system organ class and preferred term for each of the categories in the overview. Listings will be provided of all AEs, including AESIs, SAEs, TEAEs leading to discontinuation of study drug, and AEs leading to death.

3.5.2 Clinical Laboratory Tests

Chemistry, hematology, and coagulation samples will be collected at Screening, Pre-Baseline, Baseline (Day 1), Week 1, Week 4, Week 8, Week 12, Week 16 and Week 20 Safety follow-up.

All continuous laboratory parameters except of liver injury and function markers (see section 2.3.3) will be summarized descriptively by value at baseline and each planned post-baseline visit, together with the corresponding changes from baseline. All categorical laboratory parameters will be tabulated at baseline and each planned post-baseline visit.

For the selected parameters, graded treatment-emergent laboratory abnormalities will be presented for laboratory data using the grading scheme in the Common Terminology Criteria for Adverse Events (CTCAE) Version 5.0.

Laboratory	CTCAE Term	Grade 1	Grade 2	Grade 3	Grade 4
Chemistry	Alanine aminotransferase increased	>ULN - 3.0 x ULN if baseline was normal; 1.5 - 3.0 x baseline if baseline was abnormal	>3.0 - 5.0 x ULN if baseline was normal; >3.0 - 5.0 x baseline if baseline was abnormal	>5.0 - 20.0 x ULN if baseline was normal; >5.0 - 20.0 x baseline if baseline was abnormal	>20.0 x ULN if baseline was normal; >20.0 x baseline if baseline was abnormal
	Aspartate aminotransferase increased	>ULN - 3.0 x ULN if baseline was normal; 1.5 - 3.0 x baseline if baseline was abnormal	>3.0 - 5.0 x ULN if baseline was normal; >3.0 - 5.0 x baseline if baseline was abnormal	>5.0 - 20.0 x ULN if baseline was normal; >5.0 - 20.0 x baseline if baseline was abnormal	>20.0 x ULN if baseline was normal; >20.0 x baseline if baseline was abnormal
	Blood bilirubin increased	Total Bilirubin >ULN - 1.5 x ULN if baseline was normal;	Total Bilirubin >1.5 - 3.0 x ULN if baseline was normal;	Total Bilirubin >3.0 - 10.0 x ULN if baseline was normal;	Total Bilirubin >10.0 x ULN if baseline was normal;

Laboratory	CTCAE Term	Grade 1	Grade 2	Grade 3	Grade 4
		> 1.0 - 1.5 x baseline if baseline was abnormal	>1.5 - 3.0 x baseline if baseline was abnormal	>3.0 - 10.0 x baseline if baseline was abnormal	>10.0 x baseline if baseline was abnormal
	Alkaline phosphatase increased	>ULN - 2.5 x ULN if baseline was normal; 2.0 - 2.5 x baseline if baseline was abnormal	>2.5 - 5.0 x ULN if baseline was normal; >2.5 - 5.0 x baseline if baseline was abnormal	>5.0 - 20.0 x ULN if baseline was normal; >5.0 - 20.0 x baseline if baseline was abnormal	>20.0 x ULN if baseline was normal; >20.0 x baseline if baseline was abnormal
	Creatinine increased	>ULN - 1.5 x ULN	>1.5 - 3.0 x baseline; >1.5 - 3.0 x ULN	>3.0 x baseline; >3.0 - 6.0 x ULN	>6.0 x ULN
	CPK increased	>ULN - 2.5 x ULN	>2.5 x ULN - 5 x ULN	>5 x ULN - 10 x ULN	>10 x ULN
	Hyperkalemia	Potassium >ULN - 5.5 mmol/L	Potassium >5.5 - 6.0 mmol/L	Potassium >6.0 - 7.0 mmol/L	Potassium >7.0 mmol/L
	Serum amylase increased	>ULN - 1.5 x ULN	>1.5 - 2.0 x ULN	>2.0 - 5.0 x ULN	>5.0 x ULN
	Hypoglycemia	Glucose <LLN - 55 mg/dL; <LLN - 3.0 mmol/L	Glucose <55 - 40 mg/dL; <3.0 - 2.2 mmol/L	Glucose <40 - 30 mg/dL; <2.2 - 1.7 mmol/L	Glucose <30 mg/dL; <1.7 mmol/L;
	GGT increased	>ULN - 2.5 x ULN if baseline was normal; 2.0 - 2.5 x baseline if baseline was abnormal	>2.5 - 5.0 x ULN if baseline was normal; >2.5 - 5.0 x baseline if baseline was abnormal	>5.0 - 20.0 x ULN if baseline was normal; >5.0 - 20.0 x baseline if baseline was abnormal	>20.0 x ULN if baseline was normal; >20.0 x baseline if baseline was abnormal
Hematology	Anemia	Hemoglobin (Hgb) <LLN - 10.0 g/dL; <LLN - 6.2 mmol/L; <LLN - 100 g/L	Hemoglobin (Hgb) <10.0 - 8.0 g/dL; <6.2 - 4.9 mmol/L; <100 – 80 g/L	Hemoglobin (Hgb) <8.0 g/dL; <4.9 mmol/L; <80 g/L	
	Hemoglobin increased	Hemoglobin > ULN & Increase in >0 - 2 g/dL	Hemoglobin > ULN & Increase in >2 - 4 g/dL	Hemoglobin > ULN & Increase in >4 g/dL	
	CD4 lymphocytes decreased	<LLN - 500/mm3; <LLN - 0.5 x 10e9 /L	<500 - 200/mm3; <0.5 - 0.2 x 10e9 /L	<200 - 50/mm3; <0.2 x 0.05 - 10e9 /L	<50/mm3; <0.05 x 10e9 /L
	Neutrophil count decreased	<LLN - 1500/mm3; <LLN - 1.5 x 10e9 /L	<1500 - 1000/mm3; <1.5 - 1.0 x 10e9 /L	<1000 - 500/mm3; <1.0 - 0.5 x 10e9 /L	<500/mm3; <0.5 x 10e9 /L
	White blood cell decreased	<LLN - 3000/mm3; <LLN - 3.0 x 10e9 /L	<3000 - 2000/mm3; <3.0 - 2.0 x 10e9 /L	<2000 - 1000/mm3; <2.0 - 1.0 x 10e9 /L	<1000/mm3; <1.0 x 10e9 /L
	Platelet count decreased	<LLN - 75,000/mm3; <LLN - 75.0 x 10e9 /L	<75,000 - 50,000/mm3; <75.0 - 50.0 x 10e9 /L	<50,000 - 25,000/mm3; <50.0 - 25.0 x 10e9 /L	<25,000/mm3; <25.0 x 10e9 /L
Coagulation	INR increased	>1.2 - 1.5; >1 - 1.5 x baseline if on anticoagulation	>1.5 - 2.5; >1.5 - 2.5 x baseline if on anticoagulation	>2.5; >2.5 x baseline if on anticoagulation	

For the other parameters, the incidence of abnormalities (as defined by normal ranges) prior to the first dose of study drug and after the first dose of study drug will be summarized with counts and percentages of subjects.

The incidence of treatment-emergent laboratory abnormalities, defined as values that increase at least one toxicity grade from baseline at any time post baseline up to and including the date of last dose of study drugs plus 30 days, will be summarized with counts and percentages of subjects by treatment group. If baseline data are missing, then any graded abnormality (i.e., at least a Grade 1) will be considered treatment-emergent. If no pre-treatment laboratory value is available, the baseline value will be assumed to be normal (i.e., no grade [Grade 0]) for the summary of graded laboratory abnormalities.

Evaluation of drug-induced serious hepatotoxicity (eDISH) plot will be provided to display peak serum ALT and Total Bilirubin (TB) levels for each subject. The peak serum AST versus TB levels for each subject will also be provided.

3.5.3 Vital Signs

Height will be measured without shoes at Screening. Other vital signs (body weight, BMI, systolic and diastolic blood pressure, pulse, respiration rate and oral temperature) will be measured at each visit. Triplicate blood pressure and heart rate will be measured at Baseline (Day 1) and at 48 - 96 hours post-dose on Weeks 4 and Week 8.

Values and changes from baseline to each visit and timepoint for vital signs will be summarized descriptively. Triplicate measurements will be averaged prior to summary.

3.5.4 12-Lead Electrocardiograms

12-lead ECG will be measured at Screening, Baseline (Day 1) and Week 16.

Values and changes from baseline of ECG parameters (heart rate, PR interval, QRS Interval, QT interval, QTcB interval, QTcF interval and RR interval) and overall interpretation will be summarized by visit. The number and percentage of subjects with absolute QTc interval values in the pre-specified categories (>450, >480, and >500 msec), and QTc interval change (>30 and >60 msec) will be summarized.

3.5.5 Other Safety Assessments

ADA, NAB and other safety data will be listed.

3.6 Pharmacokinetic (PK) Assessment

PK sampling will be obtained pre-dose at Baseline (Day 1), Week 1, Week 4, Week 8, Week 12, Week 16, Week 20 Safety follow-up and Repeat Biopsy Week 22 - 24, and at 48-96 hours post-dose on Week 4 and Week 8.

Plasma AKR-001 concentrations will be listed and summarized by treatment group, visit, and time point for the PK Analysis Set.

3.7 Bone Biomarker Assessment

Bone biomarkers samples (Osteocalcin, P1NP, CTX-1, BSAP, PTH, and Vitamin D) will be collected at Baseline (Day 1), Week 1, Week 4, Week 8, Week 12, Week 16 and Week 20 Safety follow-up.

Values of bone biomarkers and changes from baseline to each visit will be summarized descriptively on Biomarker Analysis Set.

4 DATA MONITORING COMMITTEE (DMC)

An external DMC that consists of two hepatologists, one cardiologist and a statistician will review the progress of the study. The DMC will convene after 20 subjects (approximately 5 per treatment group) have completed the Week 4 assessments. The DMC will receive all reports of SAEs including potential DILI and liver decompensation events and convene as needed to monitor for safety. The DMC will provide recommendations to Akero whether the nature, frequency, and severity of adverse effects associated with study treatment warrant the early termination of the study in the best interests of the participants, whether the study should continue as planned, or the study should continue with modifications. The DMC may also provide recommendations as needed regarding study design.

The DMC's specific activities will be defined by a mutually agreed charter, which will define the DMC's membership, conduct, and meeting schedule.

While the DMC will be asked to advise Akero regarding future conduct of the study, including possible early study termination based on the stopping rules from the protocol, Akero retains final decision-making authority on all aspects of the study.

5 INTERIM ANALYSIS

No formal interim analysis is planned. However, a primary efficacy analysis is planned after all subjects complete Week 12 study visit, prior to the final analyses.

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

The unblinded tables will be unblinded at the treatment group level, and be limited to topline results for efficacy and safety (i.e. primary efficacy, MRI-PDFF responders, ALT, selected TEAE). Statistics that could potentially unblind at the individual level will be removed. All the details will be defined in the unblinding plan.

6 CHANGES FROM PROTOCOL-SPECIFIED STATISTICAL ANALYSES

Additional exploratory endpoints: non-HDL cholesterol, liver fibrosis responder (≥ 2 point decrease in NAS score with 0-1 point inflammation, 0 point ballooning, without worsening of fibrosis); and MRI-PDFF responders.

7 PROGRAMMING SPECIFICATIONS

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

APPENDIX A: REFERENCES

Sanyal A, Charles ED, Neuschwander-Tetri BA, Loomba R, Harrison SA, Abdelmalek MF, et al. Pegbelfermin (BMS-986036), a PEGylated fibroblast growth factor 21 analogue, in patients with non-alcoholic steatohepatitis: a randomised, double-blind, placebo-controlled, phase 2a trial. Lancet. 2019; 392(10165) :2705-2717.