Official Title of Study:

A Phase 2B Randomized, Double-Blind, Placebo-Controlled Study Evaluating the Safety and Efficacy of BMS-986036 (PEG-FGF21) in Adults with Nonalcoholic Steatohepatitis (NASH) and Compensated Liver Cirrhosis

PROTOCOL MB130-069

NCT Number: NCT03486912

Document Date (Date in which document was last revised): February 20, 2021

Statistical Analysis Plan

Sponsor:	Bristol-Myers Squibb
Protocol No:	MB130-069
[redacted]	
Protocol Version Date:	30-Jan-2020
Protocol Version No.:	4.0
Title:	A Phase 2B Randomized, Double-Blind, Placebo-Controlled Study Evaluating the Safety and Efficacy of BMS-986036 (PEG-FGF21) in Adults With Nonalcoholic Steatohepatitis (NASH) and Compensated Liver Cirrhosis
[redacted]	
SAP No.	V2.0

Approvals

Sponsor	
Sponsor Name:	Bristol-Myers-Squibb (BMS)
Representative/Title:	[redacted]
Signature/Date:	
Representative/Title:	[redacted]
Signature/Date:	
[redacted]	
Project Manager/Title:	[redacted]
Signature/Date:	
Biostatistician/Title (Author):	[redacted]
Signature /Date:	

Table of Contents

1.0 Introduction

The statistical analysis plan (SAP) describes the statistical methods to be used during the reporting and analyses of data collected under Bristol-Myers Squibb (BMS) Protocol MB130-069.

The Statistical Analysis Plan outlines the following:

- Study design
- Study objectives
- Endpoints
- Analysis populations
- Statistical methods
- Conventions and definitions

This SAP should be read in conjunction with the study protocol (version 4.0, 30 January 2020) and electronic case report (eCRF) [redacted] according to Page 1 of this document. Any further changes to the protocol or eCRF may necessitate updates to this SAP. Changes following approval of the first version SAP will be tracked in the SAP Change Log and a final version of the updated SAP will be approved prior to database lock.

2.0 Study Design

2.1 Overall Study Design

This is a multicenter, double-blind, placebo-controlled, randomized, parallel-group study to demonstrate the efficacy and safety of BMS-986036 in the treatment of participants with NASH and compensated cirrhosis.

The study will consist of 4 periods:

- A Screening Period of up to 8 weeks in duration (this period may be extended to total of up to 12 weeks upon discussion with the medical monitor)
- A 48-week, Double-Blind Treatment Period, during which the participants will receive blinded study treatment (BMS-986036 10 mg, 20 mg, 40 mg once weekly (QW) or matching placebo QW)
- A 4-week Post-Treatment Follow-up [PTFU] Period
- A Follow-Up Period of up to 14 months (1 month is defined as 4 calendar weeks) after the Week 52/PTFU Visit for the collection of an additional dual-energy X-ray absorptiometry (DXA) assessment (in all participants), samples for potential immunogenicity testing (in all participants) as well as plasma and serum for biomarkers (in all participants) at 6 months (± 14 days) Post Week 52/PTFU. Subsequent visits at 9, 12, and 14 months after Week 52 or PTFU will be performed in participants for whom Long-Term Immunogenicity Follow-Up Visits are required.

At the end of the Screening Period, participants meeting all inclusion and no exclusion criteria will enter the Double-Blind Treatment Period.

Participants will be enrolled and randomized via interactive response technology (IRT) to receive BMS-986036 10 mg QW, BMS-986036 20 mg QW, BMS-986036 40 mg QW or matching placebo QW in a 1:1:1:1 ratio. A liver biopsy will be performed during the Screening Period (if historical biopsy meeting protocol requirements is not available) and at Week 48 in all participants.

Arm 1: BMS-986036 (10 mg QW)
Arm 2: BMS-986036 (20 mg QW)
Arm 3: BMS-986036 (40 mg QW)
Arm 4: Placebo

The protocol was planned to randomize approximately 100 participants to 1 of 4 treatment groups (approximately 25 participants per group). Protocol amendment v4.0 reflects that upon close of enrollment approximately 155 participants were randomized to treatment (approximately 38 participants

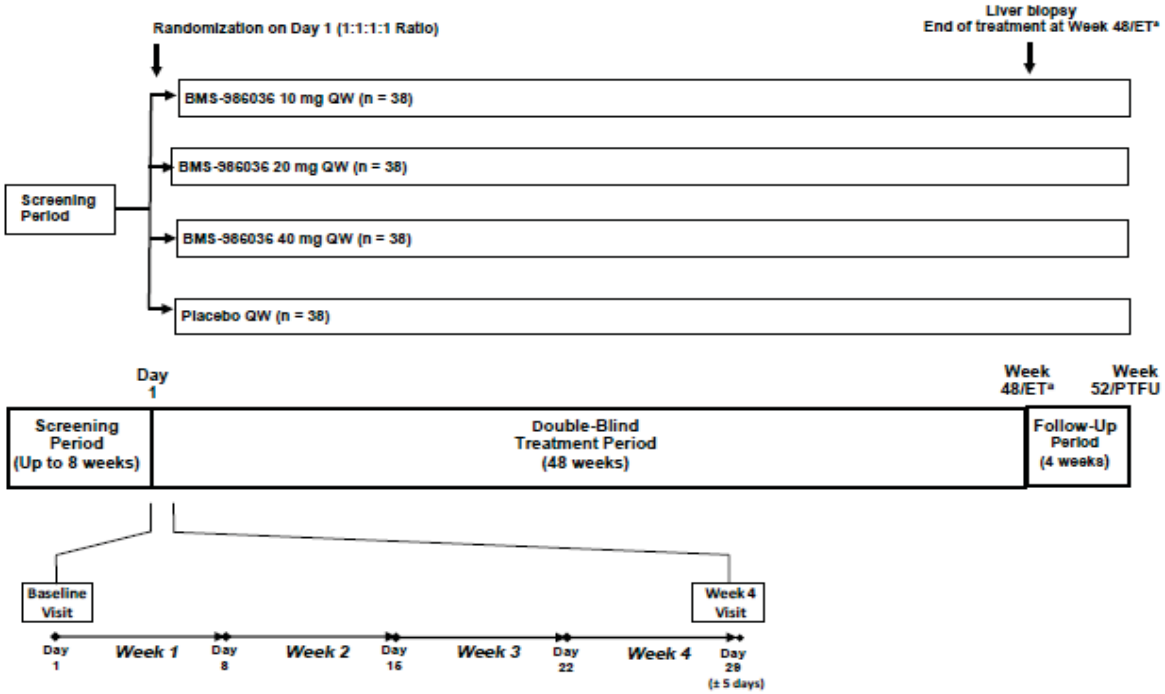
per group). The increased number of randomized participants occurred following: 1) unexpectedly high enrollment rate in the final month of the enrollment period; 2) several participants who had initially been enrolled into FALCON 1 (MB130-068) were determined (based on central pathology biopsy results) to qualify for FALCON 2 (MB130-069) during the period between the closure of enrollment and completion of randomization; and 3) an unexpected decrease in the actual screen failure rate for subjects entered in the final month of enrollment.

Based on the unmet medical need for patients, the benefit/risk assessment discussed in Protocol Section 6.3, and the strict safety monitoring of all participants, the potential clinical benefit outweighs the risks for those who have chosen to participate in this study. No additional revisions are required to the statistical considerations due to the increased number of randomized participants.

Randomized participants will receive their assigned double-blind study treatment for a total of 48 weekly doses. The treatment will be administered by subcutaneous (SC) injection of 2 prefilled syringes of study treatment in the abdomen according to their assigned treatment arm. During the Double-Blind Treatment Period, participants will be evaluated for safety and efficacy every 4 weeks (±5 days) through Week 24 and every 8 weeks (±5 days) after Week 24 through Week 48.

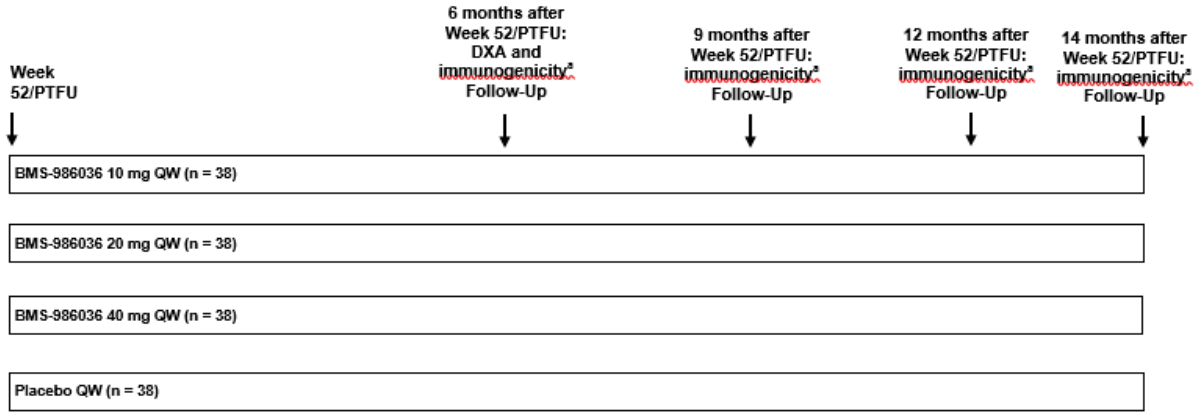
A schematic of the study design is displayed in Figure 1 (Screening to Week 52/Post-Treatment Follow-up [PTFU]) and Figure 2 (follow-up visits after the Week 52/PTFU Visit).

Figure 1. Study Design Schematic – Screening to Week 52/PTFU



ET = early termination; PTFU = Post-Treatment Follow-Up; QW = once weekly
* For participants who discontinue study medication prematurely, liver biopsy will be performed at ET if the participant has completed at least Week 20. If a participant discontinues study participation prior to Week 20 they should consider having a liver biopsy at ET.

Figure 2. Study Schematic – Bone Mineral Density and Immunogenicity Follow-Up Visits After the Week 52/PTFU Visit (visit windows ± 14 days) (1 month is defined as 4 calendar weeks)



DXA = dual-energy X-ray absorptiometry PTFU = Post-Treatment Follow-Up; QW = once weekly

[a] All participants will return 6 months (± 14 days) (1 month is defined as 4 calendar weeks) after the Week 52/PTFU Visit for dual-energy X-ray absorptiometry (DXA) to assess bone mineral density (BMD) and for collection of samples for potential immunogenicity testing as well as serum and plasma biomarker samples. Participants will return at 9 months, 12 months, and 14 months (visit windows ± 14 days) (1 month is defined as 4 calendar weeks) after the Week 52/PTFU Visit if required for immunogenicity testing (See Protocol Section 8.5.3.6).

2.2 Sample Size Considerations

The primary efficacy endpoint is the proportion of participants with a ≥1-stage improvement in fibrosis (as defined by NASH Clinical Research Network (CRN) Fibrosis Score) without worsening of NASH (as defined by the NAFLD Activity Score) at Week 48, as determined by liver biopsy.

It is planned to randomize approximately 100 participants to 1 of the 4 treatment groups (approximately 25 participants per group), BMS-986036 10 mg QW, BMS-986036 20 mg QW, BMS-986036 40 mg QW or matching placebo QW in a 1:1:1:1 ratio. The final sample size was approximately 155 participants.

2.3 Randomization

Randomization will be conducted in the following manner:

Participants will be stratified at randomization by country (Non-Japan versus Japan). Participants in countries other than Japan will be stratified at randomization according to Type 2 Diabetes Mellitus (T2DM) status (yes versus no). Japanese participants will not be further stratified.

- Eligible participants will be randomized by the IRT using a prespecified schedule in a double-blind manner to 10 mg BMS-986036, 20 mg BMS-986036, 40 mg BMS-986036, or placebo arms in a 1:1:1:1 ratio based on the stratification factors specified in Section 2.1.
- The IRT will provide study treatment kit numbers that contain the appropriate study treatment (BMS-986036 or placebo) for dispensation to that participant.

At subsequent study visits, the investigator or designee will access the IRT to receive the corresponding kit numbers assigned to the participant for the purpose of dispensing study treatment.

2.4 Unblinding Information

The Data Monitoring Committee (DMC) provides oversight of safety considerations throughout the study. A separate unblinded team, comprised of an unblinded Independent Reporting Statistician (IRS) and unblinded programmer(s), will produce output for the DMC using masked treatments. Treatment decodes may only be requested by the DMC Chair and will be provided by the IRS. Data summaries and listings will be transmitted via a secure portal by the IRS to only the DMC members. Additional details regarding the DMC process and unblinding are provided in the DMC charter.
Sponsor and designee personnel may be unblinded once all participants have completed the Week 52/PTFU Visit and all data have been collected through that time point to facilitate analyses. Designated sponsor staff may be unblinded (obtain the randomization codes) prior to database lock to facilitate the bioanalytical analysis of PK samples and immunogenicity. A bioanalytical scientist in the sponsor Bioanalytical Sciences department (or a designee in the external central bioanalytical laboratory) will be unblinded to (may obtain) the randomized treatment assignments in order to minimize unnecessary bioanalytical analysis of samples.

2.5 Clarification from Protocol

- The protocol specifies that study medication is defined as any investigational treatment(s), marketed product(s), placebo or medical device intended to be administered to a study participant according to the study randomization or treatment allocation, and study treatment includes both investigational [medicinal] product and non-investigational [medicinal] product. For the purposes of the calculations contained in this SAP, study treatment will include only doses of BMS-986036 or placebo.

[redacted]

- Impact due to COVID-19 on study visits is not described in the protocol, a summary table of patients with delayed or missed visits, and the listing of missed visits will be presented.

3.0 **Study Objectives**

3.1 Primary Efficacy Objective

The primary efficacy objective of this study is:

- To evaluate the efficacy of BMS-986036 in adults with NASH and compensated cirrhosis based on improvement in fibrosis without worsening of NASH as determined by liver biopsy at Week 48.

For this protocol, improvement in fibrosis is defined as a decrease of fibrosis by ≥ 1 point, as determined by the NASH CRN Fibrosis Score. Worsening of NASH is defined as an increase of the NAFLD Activity Score (NAS) by ≥ 1 point.

3.2 Secondary Efficacy Objectives

The secondary efficacy objectives of this study are:

- To evaluate the impact of BMS-986036 on improvement of fibrosis as determined by liver biopsy at Week 48
- To evaluate the impact of BMS-986036 on improvement in fibrosis without worsening of NASH or improvement of NASH as determined by liver biopsy at Week 48
- To evaluate the impact of BMS-986036 on resolution of NASH, as determined by liver biopsy at Week 48
- To evaluate the impact of BMS-986036 on improvement of NASH, as determined by liver biopsy at Week 48
- To evaluate the impact of BMS-986036 on collagen proportionate area (CPA) as determined by liver biopsy at Week 48

3.3 Safety Objectives

The safety objectives of this study are to demonstrate the safety of BMS-986036 in adults with NASH and compensated cirrhosis throughout the course of the study, including bone mineral density and immunogenicity.

3.4 Pharmacokinetic Objectives

- To determine trough concentrations of BMS-986036 in all participants

[redacted]

3.5 Exploratory Objectives

To explore the impact of BMS-986036 administration on the following parameters:

- Quantitative assessments of fibrosis and fat on biopsy
- Nonalcoholic fatty liver disease (NAFLD) Activity Score (NAS)
- Digital pathology assessment of NASH
- Hepatic fat fraction by magnetic resonance imaging-proton density fat fraction (MRI-PDFF)
- Noninvasive measures of fibrosis and/or cirrhosis
- Metabolic assessments
- Child-Pugh Turcotte score
- Model for End-Stage Liver Disease (MELD) score
- Liver-related clinical outcome events
- Patient-reported outcomes
- Exploratory biomarkers

4.0 Endpoints

4.1 Primary Efficacy Endpoint

The primary efficacy endpoint of this study is the proportion of participants who achieve a ≥ 1-stage improvement in fibrosis (as defined by NASH CRN Fibrosis Score) without worsening of NASH (as defined by the NAFLD Activity Score) at Week 48 as determined by liver biopsy.
Improvement in fibrosis is defined as a decrease of fibrosis by ≥ 1 point, as determined by the NASH CRN Fibrosis Score. Worsening of NASH is defined as an increase of the NAFLD Activity Score (NAS) by ≥ 1 point (details on scoring is provided in the protocol Section 8.5.1.1).

4.2 Secondary Efficacy Endpoints

The secondary efficacy endpoints are the proportion of participants with the following:

- Ishak Score >=1 stage improvement at Week 48
- NASH CRN Fibrosis Score ≥ 1-stage improvement (as defined by NASH CRN Fibrosis Score) in fibrosis without worsening of NASH or NASH improvement (reduction of NAS by ≥ 2 points with contribution from > 1 NAS component) at Week 48
- >= 1-stage improvement (as defined by NASH CRN Fibrosis Score) in fibrosis
- Any CPA decrease at Week 48
- NASH resolution (NAS component of ballooning = 0 and inflammation = 0-1) at Week 48
- NASH improvement (reduction of NAS by ≥ 2 points with contribution from > 1 NAS component) at Week 48

4.3 Safety Endpoints

Safety will be assessed by the following as per the protocol specified schedule of assessments:

- Incidence and frequency of treatment-emergent AEs and SAEs
- ECGs at Week 24, and Week 48
- Vital signs at each visit
- Physical examinations at each visit through Week 52/PTFU
- Laboratory evaluations at each visit through Week 52/PTFU
- BMD as measured by DXA Week 48/ET, and 6 months (1 month is defined as 4 calendar weeks) after the Week 52/PTFU Visit
- Anti-BMS-986036 antibodies at each visit with follow-up for up to 14 months (1 month is defined as 4 calendar weeks) after the Week 52/PTFU Visit
- Anti-FGF21 antibodies at each visit with follow-up for up to 14 months (1 month is defined as 4 calendar weeks) after the Week 52/PTFU Visit

4.3.1 Adverse Events

The definition of adverse events (AE) and serious adverse events (SAE) as well as their time periods for collection can be found in the Protocol Section 8.5.3.1 and Protocol Appendix 9.

Adverse event on special interest (AESI) will be programmatically determined by searching a predefined list of preferred terms for gastrointestinal events or bone-related events. Injection site reactions are determined via a checkbox in the eCRF. Additional details for injection site reactions will be collected including injection-site erythema and edema.

4.3.2 Electrocardiograms (ECG)

- Absolute and change from baseline values
- Post-baseline abnormalities defined by the below categories:
    - Absolute QTcF (QT interval corrected using Fridericia's formula) prolongation

- >450 msec
- >480 msec
- >500 msec

- Change from baseline in QTcF Interval
  - >30 msec
  - >60 msec

4.3.3 Vital Signs

- Absolute and change from baseline values

4.3.4 Physical Examinations

Full or abbreviated physical examinations will be conducted at each visit through Week 52/PTFU.

4.3.5 Clinical Laboratory Parameters

- Absolute and change from baseline values
- Potential hepatotoxicity by time point
  - Aspartate aminotransferase (AST) or alanine aminotransferase (ALT) > 3 x upper limit of normal (ULN) or > 5 x ULN
  - Total Bilirubin > 2 x ULN
  - (AST or ALT) >3 x ULN and Total Bilirubin > 2 x ULN
- Postbaseline abnormalities defined by the below categories:
  - Serum albumin < 2.8 g/dL
  - Platelet count < 75 x 10^3/ µL
  - Fasting plasma glucose <60 mg/dL or >350 mg/dL
  - Fasting triglycerides > 500 mg/dL
  - Estimated glomerular filtration rate (eGFR) < 30 mL/min/1.73 m^2

4.3.6 Bone Mineral Density Testing (BMD)

BMD as measured by DXA scans of the hip (including femoral neck) and spine that will be performed during the Screening Period and at the Week 48 Visit (or ET, if applicable) as indicated in the Schedule of Activities (protocol Section 1). Participants will return for a follow-up DXA assessment 6 months (1 month is defined as 4 calendar weeks) after the Week 52/PTFU Visit. DXA may be performed within a window of ± 7 days of the scheduled Week 48/ET Visit and within a window of ± 14 days of the scheduled 6-month follow-up visit. DXA scans will be evaluated by central reading. Following will be summarized.

- Absolute and change from baseline values

4.3.7 Immunogenicity

Immunogenicity (collected at Day 1 and each subsequent visit through the Week 52/PTFU visit and at long-term Immunogenicity Follow-Up visits.

- Proportion of participants with anti-BMS-986036 antibodies (including neutralizing) through Week 52/PTFU and in long-term follow-up

- Proportion of participants with anti-FGF21 antibodies (including neutralizing) through Week 52/PTFU and in long-term follow-up

4.4 Pharmacokinetic Endpoints

- Trough (predose) BMS-986036 concentrations from samples collected at each visit

[redacted]

4.5 Exploratory Endpoints

Exploratory endpoints include the following:

- Quantitative measurements of fibrosis and fat on biopsy
  - Assessment of NASH CRN Fibrosis Score at Week 48
    - Proportion of participants with >=1 stage improvement
    - Proportion of participants with >=2 stage improvement
    - Proportion of participants with >=2 stage improvement without worsening of NASH
    - Changes in Fibrosis (as expressed by Shift from baseline)
    - Proportion of participants with >=1 stage improvement, no change, or worsening
  - Assessment of Ishak Fibrosis Score at Week 48
    - Proportion of participants with >=2 stage improvement
    - Changes in Fibrosis (as expressed by Shift from baseline)
    - Proportion of participants with >=1 stage improvement, no change, or worsening
  - Assessment of qFibrosis using digital pathology at Week 48
    - Absolute, percent change, and change from baseline
  - Assessment of CPA at Week 48
    - Proportion of participants with any improvement
    - Absolute, percent change, and change from baseline
  - Assessment of fat and alpha-SMA in stained tissue by morphometry at Week 48
    - Absolute, percent change, and change from baseline
- Assessment of NAS score at Week 48
  - Changes in NAS (as expressed by shift from baseline, for each individual NAS components (Steatosis, inflammation, Ballooning) as well as Total score)
  - Proportion of participants with improvement, no change, or worsening in NAS Total
- Hepatic fat fraction as measured using MRI-PDFF at Week 24 and Week 48
  - Absolute, percent change, and change from baseline
  - Proportion of subjects in each treatment group with >=30% relative reduction, >=20% relative reduction, >=10% relative reduction, and >=5% absolute reduction
- Noninvasive measures of fibrosis and/or cirrhosis
  - Liver stiffness as determined by Magnetic Resonance Elastography (MRE) at Week 24 and Week 48
    - Absolute, percent change, and change from baseline
    - Proportion of subjects in each treatment group with >=19% or >=15% reduction
  - Fibroscan elastography results at Week 48
    - Absolute, percent change, and change from baseline
    - Proportion of subjects in each treatment group with any decrease by time point

    [redacted]
  - Noninvasive scores of hepatic fibrosis (AST-Platelet Ratio Index (APRI), Fibrosis 4 (FIB-4), Enhanced Liver Fibrosis (ELF), and NAFLD Fibrosis Score) at Week 24 and Week 48
    - Absolute, percent change, and change from baseline
    - Proportion of participants in each ELF category by time point

- Metabolic assessments at Week 24 and Week 48
  - Physical assessments (body weight, waist circumference, and body mass index (BMI)) by time point
  - Serum lipids (low-density lipoprotein (LDL), high-density lipoprotein (HDL), and triglycerides (TGs)
    - Absolute and change from baseline
    - Abnormal postbaseline TG values according to thresholds identified in SAP Section 4.3.5
  - Plasma glucose, Plasma insulin, Hemoglobin A1c
    - Absolute and change from baseline
    - Abnormal postbaseline glucose values according to thresholds identified in SAP Section 4.3.5
- Child-Pugh Turcotte Score at Week 24 and Week 48
  - Proportion of participants in each Child-Pugh Turcotte class (A, B, C) by time point
  - Shift from baseline according to class (A, B, C)
- MELD Score at each visit through Week 52/PTFU
  - Absolute and change from baseline
- The event rate for any one (first occurrence) of the liver-related clinical outcome events by Week 48
  - Proportion of participants experiencing any of the events by Week 48
  - Proportion of participants experiencing each liver-related clinical outcome event by Week 48
- Patient-reported outcomes at Week 24 and Week 48
  - 3-Level EuroQol 5 Dimension (EQ-5D-3L) Questionnaire
    - Absolute and change from baseline in visual analog score and utility index
    - Proportion of participants in each 3-level dimension by time point
  - 36-Question Short Form Quality-of-Life (SF-36) Questionnaire
    - Absolute and change from baseline in continuous variables described in SAP Section 8.0
  - Chronic Liver Disease Questionnaire-Nonalcoholic Fatty Liver Disease (CDLQ-NAFLD)
    - Absolute and change from baseline in continuous variables described in SAP Section 8.0
- Exploratory biomarkers
  - Transaminases
    - ALT normalization (ALT ≤ ULN) at Week 24 and Week 48
    - AST normalization (AST ≤ ULN) at Week 24 and Week 48
    - ALT at each visit through Week 52/PTFU
    - AST at each visit through Week 52/PTFU
    - Change from Baseline (Day 1 predose) ALT levels at each visit through Week 52/PTFU Visit
    - Change from Baseline (Day 1 predose) AST levels at each visit through Week 52/PTFU Visit.
  - PRO-C3 (N-terminal type 3 collagen propeptide)
    - Assessment of PRO-C3 levels at each visit through Week 52/PTFU
    - Change from Baseline (Day 1 predose) in PRO-C3 levels at each visit through Week 52/PTFU Visit

[redacted]

[redacted]

- Assessment of adiponectin at each visit through Week 52/PTFU
  - Absolute, percent change, and change from baseline

[redacted]

5.0 Analysis Populations

- Enrolled Population: All participants who sign informed consent.
- Randomized Population: All participants who are randomized, analyzed as per randomized treatment.
- Modified intent-to-treat (mITT) Population: All participants who are randomized and receive at least 1 dose of study medication, analyzed according to randomized treatment. All primary efficacy analyses will be conducted using this population.
- As Treated (Safety) Population: All participants who are randomized and receive at least 1 dose of study medication, analyzed according to treatment actually received. This population will be essentially the same as the mITT unless a participant received an incorrect treatment, i.e., a treatment different from what he/she was randomized to. All safety analyses will be conducted using this population.
- Pharmacokinetic (PK) Populations: All randomized participants who receive at least 1 dose of BMS-986036 and have any available concentration-time data. Available concentration-time data includes results that may be less than the lower limit of quantification (LLOQ). [redacted]

[redacted]

6.0 Statistical Methods

6.1 Efficacy Analyses

All efficacy analyses, both primary and secondary, will be performed on the Modified-Intent-to-Treat (mITT) population, unless otherwise specified, and analyzed according to randomized treatment..

6.1.1 Primary Efficacy Analysis

The primary efficacy endpoint is the proportion of participants who achieve a ≥1-stage improvement (as defined by NASH CRN Fibrosis Score) in fibrosis without worsening of NASH (as defined by NAFLD Activity Score) at Week 48, as determined by liver biopsy.

Improvement in fibrosis is defined as a decrease of fibrosis by ≥ 1 point, as determined by the NASH CRN Fibrosis Score. Worsening of NASH is defined as an increase of the NAFLD Activity Score (NAS) by ≥ 1 point.

Participants will be stratified at randomization by country (Non-Japan versus Japan). Participants in countries other than Japan will be stratified at randomization according to Type 2 Diabetes Mellitus (T2DM) status (yes versus no). Japanese participants will not be further stratified.

If lack of response in the strata prohibits analysis, the strata may be reduced to Japan/Non-Japan only or no stratification at all.

A Cochran-Armitage trend test of proportions at a one-sided 0.05 level of significance will be used to examine the linear trend among the proportions across treatment groups. The null hypothesis states that there is no association between treatment level and fibrosis stage improvement. The alternative hypothesis states that there is an association between treatment level and fibrosis stage improvement.

The proportion of participants with response will be summarized by strata as well as overall for each treatment group. Shift tables will be presented for the NAFLD Activity Score and NASH CRN Fibrosis Score. A bar chart of response/non-response in the primary endpoint will be provided.

All participants who prematurely discontinue from treatment without a biopsy at Week 48 (or ET, if applicable) or otherwise have a missing Week 48 (or ET, if applicable) biopsy result will be considered non-responders for the primary analysis of the primary efficacy endpoint.

All biopsy results will be included in the participant listings.

6.1.1.1 Sensitivity Analyses

There will be a sensitivity analysis of primary efficacy endpoint to determine the effect of the stratification variables Japan and T2DM. A Logistic regression model will be used with randomized stratification factors as main fixed effects and NASH score at baseline as well as treatment. We will have all pairwise dose comparisons against placebo and trend test with a one-sided test at alpha = 0.05 with point estimate and two-sided 90% CI will be provided. Odds ratio point estimate and two-sided 90% CI. Average across the dose groups and comparison to placebo using a nominal one-sided test at alpha=0.05 will be calculated.

A sensitivity analysis using an extended Cochran Mantel-Haenszel (CMH) correlation test will be used to assess the trend among the proportions for treatment groups with an adjustment to strata. A two-sided 0.10 level of significance will be used.

[redacted]

6.1.2 Secondary Efficacy Analyses

All secondary efficacy analyses will be conducted among the mITT population using the same methodology as the primary efficacy analysis. A completer analysis is not planned for secondary efficacy analyses.

6.1.3 Adjustment for Multiplicity

Since this is a Phase 2 study, no adjustment for multiplicity will be applied.

6.2 Safety Analyses

Safety analyses will be conducted using the safety population.

6.2.1 Adverse Events

All adverse events will be presented by treatment group and pooled BMS-986036.

A summary of treatment-emergent AEs, including the number of events reported and the number and percentage of participants reporting

- at least one AE
- any severe AE
- any treatment-related AE
- any AEs of special interest
- any severe treatment-related AE
- any AE with outcome of death

- any serious AE
- any treatment-related SAE
- any AE leading to the discontinuation of study treatment
- any treatment-related AE leading to the discontinuation of study treatment
- any AE leading to discontinuation of the study
- any treatment-related AE leading to discontinuation of the study

A breakdown of the number and percentage of participants reporting each AE, categorized by system organ class (SOC) and preferred term (PT) will be presented. Maximum severity will be summarized for all TEAE by SOC and PT. Counts for AEs will be by participant not event and participants are only counted once within each SOC or PT. AEs of special interest will be tabulated by event category and preferred term.

Only treatment-emergent AEs will be summarized in tables. All AEs will be included in listings.

6.2.2 Deaths and Serious Adverse Events

A summary of treatment-emergent SAEs, categorized by SOC and PT will be presented. Counts will be by participant, not event, and participants are only counted once within each SOC or PT. All treatment-emergent SAEs and all SAEs will be listed.

Treatment-emergent AEs with an outcome of death will be summarized by SOC and PT. All treatment-emergent AEs with an outcome of death will be listed.

6.2.3 ECGs, Vital Signs, and Physical Examinations

Based on the safety population, vital sign and 12-lead ECG parameters will be summarized by visit, treatment group, and pooled BMS-986036. All vital signs and ECG results will be listed.

ECG results will be summarized for continuous variables and will be presented for the absolute result and change from baseline. QTcF values will be presented with the implementation of corrections (i.e., Fridericia's) as defined in ICH Guidelines E14 by the following categories:

- Absolute QTcF interval prolongation:
  - QTcF interval > 450ms
  - QTcF interval > 480ms
  - QTcF interval > 500ms
- Change from baseline in QTcF interval:
  - QTcF interval increases from baseline >30ms
  - QTcF interval increases from baseline >60ms.

Physical examination results and other safety data will be listed.

6.2.4 Clinical Laboratory Parameters

Based on the safety population, continuous laboratory results will be summarized by visit, treatment group, and pooled BMS-986036 (in both SI units and US Conventional units). Categorical results such as those for urinalysis will be summarized by number and percentage of participants in each category. These summaries will be based on central laboratory results, except in rare circumstances in which only a local lab value is available then a local lab value may be used. This value will be presented in the same unit and summarized with the central results from other participants. Summaries will present both actual and change-from-baseline results for continuous values.

Altered postbaseline values will be descriptively summarized by counts and percentages of participants ever experiencing specified altered values at any postbaseline time point.

Laboratory values below the lower limit or above the upper limit of quantification will be imputed to the lower or upper limit, respectively, for summary tables. The actual values will be displayed in the listings.

Box plots will be presented by time point. All laboratory data specified in the summary tables will be present in listings.

6.2.5 Bone Mineral Density Testing (BMD)

BMD data will be summarized by visit, treatment group, and pooled BMS-986036. In rare circumstances where the number of evaluable lumbar spine segments changes over time for bone mineral density results, only participants with the same number and same segments as their baseline assessment will be included in change from baseline summaries of lumbar spine values. In addition, only participants with at least 2 evaluable lumbar spine segments at baseline will be included in the summary of the baseline time point. All bone mineral density results will be listed. Once available, the quality control (QC)-corrected values of BMD and t- and z-scores produced will be analyzed.

6.2.6 Immunogenicity

The definitions, summaries and listings described below apply to statistical analysis for anti-BMS-986036 and anti-FGF21 antibodies.

A positive BMS-986036-induced anti-BMS-986036 and anti-FGF21 antibody immunogenicity response is defined as:

1. a missing baseline immunogenicity measurement and a positive laboratory reported immunogenicity response post-baseline
2. a negative laboratory reported baseline immunogenicity response and a positive laboratory reported immunogenicity response post-baseline
3. a positive laboratory reported baseline immunogenicity response and a positive laboratory reported immunogenicity response post-baseline that has a titer value at least 4-fold greater than the baseline titer value

A persistently positive BMS-986036-induced immunogenicity response is defined as post-dose positive results at 2 or more consecutive time-points, where the first and last positive result are at least 16 weeks apart.

Immunogenicity results will be summarized as anti-drug antibody (ADA) negative or ADA positive including persistent positive for anti-BMS-936558 and/or anti-FGF21 antibodies per treatment group in the safety population. The post-baseline evaluable participants and any neutralizing results will also be included in this summary. Figures will be provided to display efficacy endpoints for ADA status. All immunogenicity results will be listed.

6.3 Pharmacokinetics

The PK summaries described in this section will be conducted among the PK population [redacted] as appropriate.

6.3.1 Serum Concentrations

Serum concentrations of BMS-986036 will be derived from trough (predose) blood samples collected from each participant before the administration of BMS-986036 dosing at each visit. Trough BMS-986036 serum concentrations (in all participants at all visits) [redacted] will be summarized by treatment and visit for both C-terminal intact and Total BMS-986036.

Serum concentrations below LLOQ will be set to missing in the computation of mean concentration values. Descriptive statistics (n, mean, standard deviation (SD), geometric mean, % coefficient variation (CV), median, min, and max) will be used to summarize the serum concentrations at each scheduled time point.

Linear and semi-logarithmic plots of the geometric mean serum concentrations + SD versus scheduled sampling times, overlaid by treatment will be provided for trough concentrations. [redacted] These plots will show study visit and will include a reference line for LLOQ on all plots with just one analyte. The plots will match the summary table results with the exception for time points with no values >LLOQ. For plotting presentation purposes, for time points with no values >LLOQ a value of one-half LLOQ will be used to plot that time point and a corresponding footnote will be added to indicate which time points were affected. All individual subject serum concentration data will be listed.

[redacted]

In addition, individual participant plots of trough concentrations will be presented for all participants and all scheduled time points among the PK population, with C-terminal Intact and total BMS-986036 concentrations on the same plot.

[redacted]

[redacted]

6.4 Exploratory Analyses

Analyses for exploratory endpoints (including biomarkers) will be performed using the mITT or safety populations. Categorical and continuous data will be summarized descriptively as presented in SAP section 6.5.

Exploratory dichotomous response variables will be summarized in a manner similar to that used for the primary and secondary efficacy analyses:

- Assessment of NASH CRN Fibrosis Score at Week 48
  - Proportion of participants with >=1 stage improvement
  - Proportion of participants with >=2 stage improvement
  - Proportion of participants with >=2 stage improvement without worsening of NASH
  - Proportion of participants with NASH resolution
  - Proportion of participants with NASH improvement
- Assessment of Ishak Fibrosis Score at Week 48
  - Proportion of participants with >=1 stage improvement
  - Proportion of participants with >=2 stage improvement
- Assessment of CPA at Week 48
  - Proportion of participants with any improvement
- Hepatic fat fraction as measured using MRI-PDFF at Week 24 and Week 48
  - Proportion of subjects with >=30% relative reduction, >=20% relative reduction, >=10% relative reduction, and >=5% absolute reduction
- Liver stiffness as determined by MRE at Week 24 and Week 48
  - Proportion of subjects with >=19% reduction
  - Proportion of subjects with >=15% reduction
- Fibroscan elastography results at Week 48
  - Proportion of subjects with any improvement

In addition, summaries of >=1 stage improvement, no change, and worsening will be tabulated and graphically presented for NASH CRN Fibrosis Score, Ishak Fibrosis Score, and NAS Total at Week 48. Participants without a Week 48 (or ET result) will be classified into the “Worse” category for these exploratory summaries.

Similarly to the primary and secondary efficacy endpoints, for exploratory endpoints based off the liver biopsy, the Week 48 result or ET biopsy will be used for Week 48 derivations. Participants without a Week 48 or ET biopsy will be classified as non-responders for liver biopsy endpoints at Week 48. Early termination visits for non-biopsy endpoints such as imaging and biomarkers may occur at other time points and will not

be summarized together with Week 48. For non-biopsy endpoints, ET will be summarized separately in the summary tables.

For continuous endpoints the analysis of covariance (ANCOVA) models will be used to explore changes from baseline that include fixed factors for the treatment group and the corresponding covariate for baseline value for the following:

- Quantitative measurements of fibrosis and fat on biopsy
  - Assessment of qFibrosis using digital pathology
  - Assessment of CPA
  - Assessment of fat and alpha-SMA by morphometry
- Hepatic fat fraction as measured using MRI-PDFF
- Noninvasive measures of fibrosis and/or cirrhosis
  - Liver stiffness as determined by MRE
  - Fibroscan elastography results [redacted]
  - APRI, FIB-4, and NAFLD scores
- Exploratory biomarkers
  - Transaminases ALT and AST
  - PRO-C3

A repeated measures analysis will model changes from baseline for each of the exploratory endpoints separately for data collected at planned visits using restricted maximum likelihood estimations. The model will include treatment group, visit, baseline value, treatment * visit and baseline stratification factors as fixed effects. An unstructured covariance matrix will be used to represent the correlation of the repeated measures within each subject. This model will be used to provide point estimates for the treatment differences, standard errors for the differences, and corresponding 2-sided 95% confidence intervals for mean change from baseline at each time point. Statistical inferences for comparing treatment groups will utilize Kenward-Rodgers Degrees of Freedom.

Other biomarkers will be descriptively summarized by time point and will include percent change from baseline in addition to change from baseline.

6.5 General Methodology

All analyses will use SAS® version 9.4 or higher.

- Descriptive summaries will be tabulated by treatment group and overall, unless otherwise specified.
- Categorical data will be presented using counts and percentages, with the number of participants in each category as the denominator for percentages
  - For by-visit summaries denominators may be adjusted to reflect the number of participants in the study at the time of the visit
  - Percentages will be rounded to one decimal place except 0% and 100% which will be displayed without any decimal places and percentages will not be displayed for zero counts.
- Continuous data will be summarized using the number of observations (n), mean, SD, median, first quartile, third quartile, minimum, and maximum. Minimum and maximum will be rounded to the precision of the original value.
  - Mean, median, first and third quartiles will be rounded to 1 decimal place greater than the precision of the original value. The SD will be rounded to 2 decimal places greater than the precision of the original value, with the original value having up to a maximum of 3 decimal places.

- P-values will be presented to 3 decimal places.

6.5.1 Participant Disposition

The number and percentage of participants in each analysis population will be presented. The number and percentage of participants randomized but not treated will be included.

The number of participants randomized will be presented by treatment group, along with the number and percentage of participants who completed, discontinued, and are ongoing treatment or the study at the time of the data cut by treatment group.

Enrollment by country and center will be tabulated among randomized participants.

Also, by-participant listings will be generated.

6.5.2 Demographics and Baseline Disease Characteristics

Demographics and baseline disease characteristics will be summarized by treatment group and overall for the mITT population. Individual participant listings will be provided to support the summary tables.

6.5.3 Medical History

A summary of general medical history will be presented by Medical Dictionary for Regulatory Activities (MedDRA) Version 20.1 or higher SOC and PT for the safety population. The number and percentage of participants with each event will be presented by SOC and PT. Note that counting will be by participant not event and participants are only counted once within each SOC and PT.

A separate specific disease history table, including history of prior NASH therapy and history of Hepatitis C virus, will be provided by treatment group and total for the safety population. The proportion of participants with each type of disease history condition (e.g. history of type 1 diabetes, history of type 2 diabetes, etc.) will be included. These histories will be presented in participant listings.

Data collected on prior surgeries will be listed.

6.5.4 Prior Medications

Prior medications summaries will be presented for the safety population by treatment group and overall. A summary showing the number and percentage of participants who had prior medications will be presented by World Health Organization (WHO) Anatomic Therapeutic Classification (ATC) level and preferred name, version WHO DD 2017 SEP01 DDE+HD or higher. Note that counting will by participant and participants are only counted once within each ATC and PT.

6.5.5 Protocol Deviations

Important protocol deviations will be summarized by number and percentage of participants experiencing each deviation category. All protocol deviations will be listed.

6.5.6 Treatments

6.5.6.1 Extent of Study Treatment Exposure

Summary statistics for study exposure will be based on the safety population and presented by treatment group and pooled BMS-986036.

The following summaries will be included:

- Number of participants receiving at least 1 injection
- Number of participants with at least 1 injection not taken as planned and reason
- Frequency of the number of weeks of treatment (e.g. 16 weeks, 20 weeks, etc.)

- Continuous summary of treatment duration (in weeks)
- Summary of percentage of treatment completed

Participant listings will present exposure data at the injection-level. End of treatment reasons will be listed separately, as well as device malfunctions.

6.5.6.2 Concomitant and Follow-up Medications

Concomitant medications (through Week 52 / PTFU), bone-density related medications and NASH therapies taken after the end of study treatment (Week 52/PTFU through 6-month Follow-up) will be presented separately among the safety population by treatment group and overall. The summaries will include the number and percentage of participants who had medications and will be presented by WHO ATC drug name and preferred name, version WHODD March 2015 or later. A participant listing of medications will be provided to support the tables.

6.5.6.3 Diagnostic and Medical Procedures

Diagnostic and medical procedures will be presented in a participant listing.

6.5.7 Delayed or Missed Study Visits Due to COVID-19

Any disruptions to study visits (resulting in missed or delayed visit) due to COVID-19 occurred during treatment period and post treatment follow-up will be summarized using mITT population. Patients with missed or delayed visits will be tabulated using frequency and percentage by study visit. A listing of patients reporting missed study visits due to COVID-19 will be presented.

7.0 Sequence of Planned Analyses

Locking, freezing, cleaning, and details for transferring of data will be described in a separate study plan.

7.1 Week 52 Analysis

The Week 52 data analysis is based on all subjects who have completed the Week 52/PTFU visit. A data cut-off date of 08Oct2020 is applied to the raw data extract for the locked database. The study will continue to collect data for the long-term follow-up. in a blinded fashion.

- All Efficacy data will be summarized based on mITT principle for DB Lock defined by eCRF date cut-off 08Oct2020 on all participants who have completed liver biopsy.
- If the biopsies were delayed (due to impact of COVID-19 or other causes) and performed between Week 48 and Week 52, the biopsies results will be included in the planned Week 48 visit.
- Safety data will be summarized using Safety population through DB lock defined by eCRF date cut-off date of 08Oct2020.
- All available data beyond Week 52 will be summarized descriptively.

Sponsor and designee personnel may be unblinded once all participants have completed the Week 52/PTFU visit and all data have been collected through that time point to facilitate analyses. The list of TFLs to be provided for this analysis will be kept in a separate document outside the SAP.

7.2 Long-term Follow-up Analysis

Data collected after the Week 52/PTFU visit (or early termination) will be reported in a separate analysis once all participants have completed long term follow-up. The list of TFLs to be provided for this analysis will be kept in a separate document outside the SAP.

7.3 Exploratory Analyses

Any exploratory analyses completed to support study analyses in the Clinical Study Report (CSR), which were not identified in the statistical analysis plan, will be documented as such in the CSR.

8.0 **Conventions and General Definitions**

8.1 General Definitions

Term	**Definition**
Study Day	Study day is calculated as: assessment date – date of first dose + 1
Baseline	Unless otherwise stated, Baseline is defined as the last measurement prior to dosing on Day 1 (Week 0). If the measurement on Day 1 is missing or not available, then a prior measurement during the screening period may be used as baseline.
Change from Baseline	Change from baseline is defined as (value at post-baseline visit – value at baseline).
Change in the maximum post-baseline value	Change from baseline in the maximum post-baseline value is defined as highest observed value post-baseline. The change is calculated using this value as the post-baseline value.
Percent Change from Baseline	Percent change from baseline is defined as ([value at post-baseline visit – value at baseline]/value at baseline) x 100. If the baseline value is 0 and the post-baseline value is also 0, then the percent change from baseline is set to 0. Percent change from baseline is not calculated if the baseline value is zero or baseline or post-baseline value is missing.
Study Treatment	For the purposes of the calculations included in this SAP, study treatment will include only doses of BMS-986036 (dose levels of 10 mg QW, 20 mg QW or 40 mg QW) or placebo QW.
Duration of Study Treatment (weeks)	Duration of study treatment in weeks will be defined as: (date of last dose of study treatment – date of first dose of study treatment+7)/7.
Treatment Completion (%)	Treatment completion (%) will be calculated as the total number of weeks of treatment received divided by the total number of expected doses (48), multiplied by 100.
End of Study (EOS) Date	The EOS date is the date recorded on the eCRF that a participant either discontinued or completed the study. If the participant is lost

	to follow-up, the EOS date will be the date of the last visit assessment obtained.
First Dose Date – Study	The date a participant received their first dose as recorded in the eCRF as date study treatment was administered.
Last Dose Date – Study	The date of last recorded dose on the eCRF for a subject.
Unscheduled assessments	Unscheduled measurements will be included in the listings. With the exception of unscheduled measurements used for baseline and as specified in section 8.12, unscheduled measurements will be excluded from the descriptive statistics and statistical analysis.
Prior, Concomitant, and Follow-up Medications	Prior medications are defined as medications taken prior to the first dose of study treatment and discontinued prior to the first dose of study treatment. Concomitant medications are defined as any medication started before/on/after the first dose of study medication until the Week 52/PTFU Visit. Follow-up medications are defined as any medication taken after Week 52/PTFU Visit until the 6 Month Post PTFU Visit that could potentially impact bone mineral density assessments. Any Concomitant medications continued Post Week 52/PTFU will also be considered as Follow-up medication.

8.2 Missing, Unknown, or Partial Dates Imputation

Start Date		Stop Date: Complete: *yyyymmdd* <1st dose	Complete: *yyyymmdd* ≥1st dose	Partial: *yyyymm* <1st dose *yyyymm*	Partial: *yyyymm* ≥1st dose *yyyymm*	Partial: *yyyy* <1st dose *yyyy*	Partial: *yyyy* ≥1st dose *yyyy*	Missing/ Ongoing
Partial: *yyyymm*	= 1st dose *yyyymm*	2	1	n/a	1	n/a	1	1
	≠ 1st dose *yyyymm*		2	2	2	2	2	2
Partial: *yyyy*	= 1st dose *yyyy*	3	1	3	1	n/a	1	1
	≠ 1st dose *yyyy*		3		3	3	3	3
Missing		4	1	4	1	4	1	1

1 = Impute as the date of first dose
2 = Impute as the first of the month
3 = Impute as January 1 of the year
4 = Impute as January 1 of the stop year
Note: If the start date imputation leads to a start date that is after the stop date, then there is a data error and the start date is not imputed.

Imputation rules for partial or missing stop dates:

1. Initial imputation
    a. For partial stop date "mmyyyy", impute the last of the month.
    b. For partial stop date "yyyy", impute December 31 of the year.
    c. For completely missing stop date, do not impute.
2. If the stop date imputation leads to a stop date that is after the death date, then impute the stop date as the death date.
3. If the stop date imputation leads to a stop date that is before the start date, then there is a data error and the stop date is not imputed.

Imputation rules for partial or missing death dates:

1. If death year and month are available but day is missing:
    a. If "mmyyyy" for last contact date = "mmyyyy" for death date, set death date to the day after the last contact date.
    b. If "mmyyyy" for last contact date < "mmyyyy" for death date, set death date to the first day of the death month.
    c. If "mmyyyy" for last contact date > "mmyyyy" for death date, data error and do not impute.

If both month and day are missing for death date or a death date is totally missing, set death date to the day after the last contact date.

8.3 Visit Windows

Visit windows as defined in the protocol are closely monitored by clinical monitors. A protocol deviation will be documented if assessments were recorded outside the visit windows, but the assessment is recorded in the clinical database at the planned visit. For longitudinal summaries of data, assessment recorded at the planned visits in the clinical database are included and displayed in tables with visit based structure.

If there are multiple records within the same visit window, then the value in the visit window closest to the day of the planned visit is selected.

8.4 Treatment-Emergent Adverse Events

Treatment-emergent adverse events (TEAEs) are those that first occur or increase in severity or relationship to study treatment after the first dose of study treatment and not more than 30 days after the last dose of study treatment. All AEs that change in severity or relationship to study treatment are assigned a new start date and captured as a new record.

The definitions of AE and SAE can be found in the Protocol Appendix 9. If an AE causality or severity, seriousness or relationship to IP is missing, they will not be imputed.

8.5 Adverse Events of Special Interest

A list of adverse events of special interest (AESI) for BMS-986036 based on the known biologic class effects, the mechanism of action, and clinical study data are:

- Injection site reactions
- Gastrointestinal events
    - Diarrhea, frequent bowel movements, nausea, vomiting, abdominal pain
- Bone-related events

Osteoporosis, osteopenia, bone and joint injuries, fractures (except tooth fracture), endocrine abnormalities of gonadal function not elsewhere classified (NEC), hyperparathyroid disorders, hypoparathyroid disorders, parathyroid disorders NEC, parathyroid analyses, metabolic bone disorders, vitamin D abnormal, vitamin D decreased, vitamin D deficiency, and miscellaneous events related to bone density

8.6 Liver Biopsy Scoring

Histological assessment of the liver tissue sample will be conducted by a blinded Central Pathologist. The central reader will be a medical doctor, board certified in pathology, with experience in liver pathology in a clinical study setting. Additional details on the acquisition, quality requirements, histological preparation, and shipping of histological samples are in the Central Laboratory Manual and the Central Pathology Manual or Charter.

Histological Scoring

NASH CRN

For associated primary and secondary endpoints, the NASH CRN system will be used to score the histologic samples and results are reported as the NAFLD Activity Score (NAS) and the NASH CRN Fibrosis Score. The NASH CRN system is based on the concept that necroinflammatory lesions and stage of fibrosis should be evaluated separately; it assesses liver biopsies for degree of steatosis (0-3), lobular inflammation (0-3), hepatocellular ballooning (0-2), and fibrosis (0-4).

NAS

The 3 categories of steatosis, lobular inflammation, and ballooning scores are added together in an unweighted fashion to determine the NAS, which ranges from 0 to 8 (see below).

Histology Variable	Grade	Score
Steatosis	< 5%	0
	5%-33%	1
	> 33% - 66%	2
	> 66%	3
Lobular Inflammation[a]	none	0
	< 2	1
	2 - 4	2
	> 4	3
Ballooning	none	0
	few	1
	many	2

NAS = NAFLD Activity Score
[a] Foci per x200 field

Fibrosis

NASH CRN Fibrosis

Fibrosis is staged separately from NAS on a 0-4 scale: 0 (none); 1 (perisinusoidal or periportal fibrosis); 2 (perisinusoidal and portal/periportal fibrosis); 3 (bridging fibrosis); 4 (cirrhosis).

Ishak

A modified Ishak scoring system will also be used to stage fibrosis in the histologic samples. The Ishak system (0-6 scale) was originally developed to grade portal-based liver fibrosis associated with viral

hepatitis. The modified Ishak system has been adapted to stage central-based liver fibrosis associated with NASH, and it also uses a 0-6 scale:

- 0: No fibrosis
- 1: centrilobular pericellular fibrosis
- 2: centrilobular and periportal fibrosis
- 3: bridging fibrosis (few bridges)
- 4: bridging fibrosis (many bridges)
- 5: early or incomplete cirrhosis
- 6: established or advanced cirrhosis

8.7 Child-Pugh Turcotte Score

The Child-Pugh Turcotte Score assesses the severity of cirrhosis and has been shown to be an accurate measure across a broad spectrum of liver disease. It employs numerical scores of 5 measures of liver disease: total bilirubin, serum albumin, International Normalized Ratio (INR), ascites, and encephalopathy. The sum of scores from each component is the final score (see below) used for determination of class.

Parameter	Classification	Score
Bilirubin (Total)	< 2 mg/dL (< 34.2 µmol/L)	+1
	2-3 mg/dL (34.2-51.3 µmol/L)	+2
	> 3 mg/dL (> 51.3 µmol/L)	+3
Albumin	> 3.5 g/dL (> 35 g/L)	+1
	2.8-3.5 g/dL (28-35 g/L)	+2
	< 2.8 g/dL (< 28 g/L)	+3
International Normalized Ratio	< 1.7	+1
	1.7-2.2	+2
	> 2.2	+3
Ascites	Absent	+1
	Slight	+2
	Moderate	+3
Encephalopathy	No encephalopathy	+1
	Grade 1-2	+2
	Grade 3-4	+3

Class A: 5 to 6 points (least severe liver disease)

Class B: 7 to 9 points (moderately severe liver disease)

Class C: 10 to 15 points (most severe liver disease)

8.8 Liver-Related Clinical Outcome Events

Participants will be evaluated for the following liver-related clinical outcome events at each visit from Day 1 through Week 52/PTFU:

- All-cause mortality
- MELD Score ≥ 15 (with at least a 2-point increase from Baseline)
- Liver transplant
- Ascites requiring medical intervention
- Hospitalization (≥ 24 hours) for onset of variceal bleed
- Hospitalization (≥ 24 hours) for hepatic encephalopathy
- Hospitalization (≥ 24 hours) for spontaneous bacterial peritonitis
- Hepatocellular Carcinoma (HCC)

Liver-related clinical outcome events will also be reported as AEs/SAEs, as applicable, as per (Protocol Section 8.5.3.1 and Appendix 9).

8.9 CLDQ-NAFLD Scoring

The CLDQ-NAFLD questionnaire (Protocol Appendix 7) is a disease-specific quality-of-life instrument for nonalcoholic fatty liver disease and NASH. The CLDQ-NAFLD consists of 36 items in 6 domains: abdominal symptoms, activity, emotional function, fatigue, systemic symptoms, and worry. Each item is on a Likert scale with 1 representing the most impairment and 7 representing the least impairment.

Each domain score is an average of its constituent items, and the overall score is an average of the six domains. Any out-of-range value will be considered missing.

In case of missing values, calculation of the domain scores will only be derived when at least 50% of the items are completed, and the total score will require at least three domains.

Items and domains: Ni refers to score obtained for Question i.

Abdominal symptoms (AS) = Mean of N1+N5+N17

Activity/Energy (AE) = Mean of N7 + N9 + N14 + N30+ N31

Emotion (EM) = Mean of N10 + N12 + N15 + N16 + N19 + N20 + N24 + N26 + N34

Fatigue = Mean of N2 + N4 + N8 + N11 + N13 + N35

Systemic (SY) = Mean of N3 + N6 + N21 + N23 + N27 + N36

Worry (WO) = Mean of N18 + N22 + N25 + N28 + N29 + N32 + N33

Overall = Mean of six domains

8.10 EQ-5D-3L Scoring

The EQ-5D-3L questionnaire (Protocol Appendix 7) is a general quality-of-life instrument that consists of 2 components, health state description and evaluation. The health state description component measure mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Participants will rate their level of impairment in each dimension using a 3-level scale. In the evaluation component, participants evaluate their overall health status using a visual analog scale. Any out-of-range value will be considered missing.

The [redacted] programming team will also calculate a United States population-based utility index according to Shaw et al. 2005, using a validated program provided by BMS. The index score will not be generated when responses are missing for one or more of the five dimensions.

8.11 SF-36 (Version 2)

The SF-36 (Protocol Appendix 6) is a 36-item instrument. Items are divided into eight concepts plus one health comparison question.

The [redacted] programming team will use [redacted] software [redacted] to score the SF-36.

The raw scale scores are computed by summing item responses within each concept after recoding the individual item if needed.

The eight concepts to be scored are:

Physical Functioning (PF) (items 3a-3j): The ten items are coded as 1-3. There is no recoding needed. The lowest possible raw score is 10, the highest is 30 and the range is 20. High score indicates better PF.

Role-Physical (RP) (items 4a-4d): The items are coded as 1-5. There is no recoding needed. The lowest possible raw score is 4, the highest is 20 and the range is 16. High score indicates better Role- Physical function.

Role-Emotional (RE) (items 5a-5c): The items are coded as 1-5. There is no recoding needed. The lowest possible raw score is 3, the highest is 15 and the range is 12. High score indicates better Role-Emotional function.

Social Functioning (SF) (items 6, 10): The items are coded as 1-5. Item 6 is recoded as 5-1. Item 10 does not require recoding value. The lowest possible raw score is 2, the highest is 10 and the range is 8. High score indicates better social functioning.

Bodily Pain (BP) (items 7 & 8): This concept requires special coding. Item 7 is coded as 1-6 and is recoded as 6.0, 5.4, 4.2, 3.1, 2.2, 1.0. Item 8 is coded as 1-5 and is recoded as 6-1 if both item 7 and item 8 are answered. If item 7 is not answered, item 8 is recoded as 6.0, 4.75, 3.5, 2.25, 1.0. The lowest possible raw score is 2, the highest is 12 and the range is 10. High score indicates lack of bodily pain.

Mental Health (MH) (items 9b, 9c, 9d, 9f & 9h). The items are coded as 1-5. Items 9b, 9c and 9f do not require recoding. Items 9d and 9h are recoded as 5-1. The lowest possible raw score is 5, the highest is 25 and the range is 20. High score indicates better mental health.

Vitality (VT) (items 9a, 9e, 9g, 9i): The items are coded as 1-5. Items 9a and 9e are recoded as 5-1. Items 9g and 9i do not require recoding. The lowest possible raw score is 4, the highest is 20, the range is 16. High score indicates more vitality.

General Health (GH) (items 1, 11a-11d): The items are coded as 1-5. Items 11 a and 11 c do not require recoding. Items 11b and 11d are recoded as 5-1. Item 1 is recoded as 5.0, 4.4, 3.4, 2.0, 1.0. The lowest possible raw score is 5, the highest is 25 and the range is 20. High score indicates better general health perceptions.

Reported Health Transition (HT) (item 2). Item 2 is coded as 1-5 and is recoded 5-1.

Any out-of-range value will be considered missing.

Missing or blank values will be estimated using the average recoded score for each concept.

The formula for the transformed scale = [(actual raw score - lowest possible raw score)/Possible raw score range] x 100. The raw scores will be transformed to a 0 – 100 scale.

In addition to the eight concepts above, the scoring software also provides results for two aggregate scores: Physical Component Summary (PCS) and Mental Component Summary (MCS).

8.12 Fibroscan

A minimum of 10 valid readings, with at least a 60% success rate of all measurements taken

and an interquartile range of less than or equal to 30% of the median value is required for values to be included in summary tables. All values will be listed.

8.13 Other Calculations

- ELF assessment combines hyaluronic acid, procollagen 3 amino terminal peptide, and TIMP-1 (tissue inhibitor of metalloproteinases 1). A proprietary algorithm is used to evaluate each of these markers by immunoassay, to create an ELF Score which is provided by an external vendor.
- FIB4 Score = (age [years] × AST level [U/L])/ (platelet count [x10^9/L] × square root of ALT [U/L]).
- APRI Score = ([AST divided by AST Upper Limit of Normal]/platelet count [× 10^9/L]) × 100
- NAFLD Fibrosis Score = -1.675 + (0.037 × age [years] + (0.094 × BMI [(kg/m2)])+ (1.13 × Impaired Fasting Glucose or diabetes [yes = 1, no = 0]) + (0.99 × AST/ALT ratio) – (0.013 × platelet count [× 109/L]) – (0.66 × albumin [g/dL]). Note, the status of impaired fasting glucose or diabetes (yes or no) will be assessed only once at baseline as recorded by the site in the Specific Disease History eCRF page. This value will be used for all subsequent calculations.
- MELD Score = (9.57*ln[creatinine]) + (3.78 * ln[Bilirubin]) + (11.20* ln[INR]) + 6.43

Note for FIB4, APRI, NAFLD Fibrosis, and MELD scores that if any one component is missing at a given time point, another non-missing value within an absolute range of 3 days may be used. If there is more than one value that is within 3 days, the closest value after the visit date will be selected. Otherwise, if all components are not available the score will not be calculated at that time point.

9.0 **References**

SAS Institute Inc. 2015. Base SAS 9.4 Procedures Guide. 5th ed. Cary, NC: SAS Institute Inc.

Shaw JW, Johnson JA, Coons SJ. US Valuation of the EQ-5D Health States Development and Testing of the D1 Valuation Model. Medical Care. Vol 43(3): March 2005.

Younossi ZM, Stepanova M, Henry L, Racila A, Lam B, Pham HT, Hunt S. A disease-specific quality of life instrument for non-alcoholic fatty liver disease and non-alcoholic steatohepatitis: CLDQ-NAFLD. Liver Int. 2017 Aug;37(8):1209-1218. doi: 10.1111/liv.13391. Epub 2017 Mar 13.

Appendix 1 Glossary of Abbreviations

Glossary of Abbreviations:	
[redacted]	
ADA	anti-drug antibody
AE	adverse event
AESI	Adverse event of special interest
ALT	alanine aminotransferase
ANCOVA	Analysis of covariance
APRI	AST-to-platelet ratio index
AST	aspartate aminotransferase
AUC	area under the concentration-time curve
ATC	Anatomic Therapeutic Classification
[redacted]	
BMD	Bone mineral density
BMI	body mass index
BMS	Bristol-Myers Squibb
[redacted]	
CSR	Clinical study report
CI	confidence interval
CMH	Cochran Mantel-Haenszel
CPA	collagen proportionate area
CLDQ-NFLD	Chronic Liver Disease Questionnaire-Nonalcoholic Fatty Liver Disease
CSR	Clinical Study Report
CRN	Clinical Research Network
[redacted]	
CV	coefficient of variation
DMC	data monitoring committee
DXA	dual-energy X-ray absorptiometry
eCRF	Electronic Case Report Form
ECG	Electrocardiogram
eGFR	Estimated glomerular filtration rate
ELF	enhanced liver fibrosis
EOS	end of study
EQ-5D-3L	3-Level EuroQol 5 Dimension (quality-of-life questionnaire)
[redacted]	

ET	early termination
FGF	fibroblast growth factor
FGF21	fibroblast growth factor 21
FIB4	Fibrosis 4
HCC	hepatocellular carcinoma
HDL	high-density lipoprotein
INR	international normalized ratio
IRS	independent reporting statistician
IRT	interactive response technology
LDL	low-density lipoprotein
LLOQ	Lower level of quantification
[redacted]	
MedDRA	Medical Dictionary for Regulatory Activities
MELD	Model for End-Stage Liver Disease
mITT	modified intent-to-treat
MRE	magnetic resonance elastography
MRI	magnetic resonance imaging
MRI-PDFF	magnetic resonance imaging-proton density fat fraction
NAFLD	Nonalcoholic fatty liver disease
NAS	NAFLD Activity Score
NASH	nonalcoholic steatohepatitis
[redacted]	
PD	pharmacodynamic
PDFF	proton density fat fraction
PEG	polyethylene glycol
PK	pharmacokinetic
PRO-C3	N-terminal type 3 collagen propeptide
PT	Preferred term
PTFU	Post-Treatment Follow-Up
QC	quality control
QTcF	QT interval corrected using Fridericia's formula
QW	once weekly
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	standard deviation

SC	subcutaneous
SMA	smooth muscle actin
SOC	system organ class
SF-36	36-question Short Form quality-of-life questionnaire (Version 2)
T2DM	type 2 diabetes mellitus
TEAE	treatment-emergent adverse event
TFLs	Tables, figures, and listings
TG	triglycerides
TIMP-1	tissue inhibitor of metalloproteinases 1
ULN	upper limit of normal
[redacted]	
WHO	World Health Organization
[redacted]	

Appendix 2 Tables, Figures, Listings, and Supportive SAS Output Appendices

Refer to the study TFL shell document.

Statistical appendices will be provided for primary and secondary efficacy endpoints containing raw SAS output.