STATISTICAL ANALYSIS PLAN

Protocol Title:	A Multicenter, Randomized, Double-Blind, Placebo Controlled, Parallel-Group Study of NST-4016 in Patients With Nonalcoholic Steatohepatitis (NASH)
Protocol Number:	NST-02
Protocol Version/Date:	Version 8.0 / 21 July 2022
Investigational Product:	NST-4016 Icosabutate
Sponsor:	NorthSea Therapeutics B.V. Gooimeer 2-35 1411 DC Naarden The Netherlands
SAP Version/Date:	Version 1.0 / 14 February 2023

SIGNATURE PAGE

Protocol Title: A Multicenter, Randomized, Double-Blind, Placebo Controlled, Parallel-Group Study of NST-4016 in Patients With Nonalcoholic Steatohepatitis (NASH)

Protocol Number: NST-02

SAP Version/Date: Version 1.0 / 14 February, 2023

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

Signature		Date
Suzanne Lloyd	*Electronically signed by: Suzanne Lloyd Reason: Approved Date: Feb 14, 2023 16:37 GMT*	14-Feb-2023
Suzanne Lloyd Manager, Statistical Programming Medpace, Inc.		
Ping Xu	*Electronically signed by: Ping Xu Reason: Approved Date: Feb 14, 2023 12:59 EST*	14-Feb-2023
Ping Xu Director, Biostatistics Medpace, Inc.		
Xinyan Qu	*Electronically signed by: Xinyan Qu Reason: Approved Date: Feb 14, 2023 11:51 EST*	14-Feb-2023
Xinyan Qu,PhD Clinical Pharmacologist Medpace, Inc		
Holly Schachner, MD	*Electronically signed by: Holly Schachner, MD Reason: Approved Date: Feb 14. 2023 12:31 EST*	14-Feb-2023
Holly Schachner, MD Head of Clinical Development NorthSea Therapeutics B.V.		
	Electronically signed by: Stephen Harrison Reason: Approved Date: Feb 14, 2023 10:50 CST	14-Feb-2023
Stephen Harrison, MD Chief Medical Officer NorthSea Therapeutics B.V.		

David Fraser

Electronically signed by: David Fraser
Reason: Approved
Date: Feb 15, 2023 17:53 GMT+1

15-Feb-2023

David Fraser
Chief Scientific Officer
NorthSea Therapeutics B.V.

VERSION HISTORY

Version	Version Date	Description
1.0	14 February 2023	First version

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation	Definition
ATC	Anatomical Therapeutic Chemical
ADaM	Analysis Data Model
AE	Adverse event
AESI	Adverse event of special interest
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
ANCOVA	Analysis of Covariance
AST	Aspartate aminotransferase
AUC	Area under the concentration time curve
AUC_{0-t}	Area under the plasma concentration-time curve from time 0 to the time of the last observed concentration
AUC_{0-inf}	Are under the plasma concentration-time curve from time 0 extrapolated to infinity
BMI	Body Mass Index
CDISC	Clinical Data Interchange Standards Consortium
CL/F	Apparent plasma clearance
C_{max}	Maximum observed plasma concentration
COVID-19	Coronavirus Disease 2019
CRF	Case report form
CSR	Clinical Study Report
cT1	Iron-corrected T1
DSMC	Data Safety Monitoring Committee
ECG	Electrocardiogram
eCRF	Electronic case report form
ELF	Enhanced liver fibrosis
ET	Early Termination
GGT	Gamma glutamyl transferase
HOMA-IR	Homeostatic Model Assessment for Insulin Resistance
hsCRP	High-sensitivity C-reactive protein
ITT	Intent-to-Treat
LS	Least square
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified Intent-to-Treat
MMRM	Mixed-Model Repeated Measures
MNAR	Missing not a random
MRI	Magnetic resonance imaging
MRI-PDFF	Magnetic resonance imaging-proton density-fat fraction
NAS	Nonalcoholic Fatty Liver Disease Activity Score
NASH	Nonalcoholic steatohepatitis
PK	Pharmacokinetic(s)
PP	Per-Protocol
Pro-C3	N-terminal type 3 collagen propeptide

Abbreviation	Definition
SAE	Serious adverse event
SAP	Statistical Analysis Plan
STDM	Standard Data Tabulation Model
TEAE	Treatment-emergent adverse event
TEAESI	Treatment-emergent adverse event of special interest
TESAE	Treatment-emergent serious adverse event
TFL	Tables, Figures and Listings
T_{max}	Time to reach the maximum observed plasma concentration
WHO	World Health Organization
V_z/F	Apparent volume of distribution

1 INTRODUCTION

The purpose of this Statistical Analysis Plan (SAP) is to provide a description of the statistical methods to be implemented for the analysis of data from the study with protocol number NST-02. The SAP will be finalized prior to database lock. Any deviations from the SAP after database lock will be documented in the final Clinical Study Report (CSR).

2 STUDY OVERVIEW

2.1 Study Objectives

2.1.1 Primary Objective

The primary objective is to evaluate the efficacy of different doses of NST-4016 on the resolution of nonalcoholic steatohepatitis (NASH) without worsening of fibrosis in patients with NASH without histological or clinical evidence of cirrhosis.

2.1.2 Secondary Objective

The secondary objectives of the study are:

- To determine the safety and tolerability of NST-4016 in patients with NASH
- To characterize the pharmacokinetics (PK) of NST-4016

2.2 Study Design

2.2.1 Overview

This was a 52-week (62 week, including screening and follow-up), multicenter, randomized, double-blind, placebo-controlled, parallel-group study in male and female patients with a histological diagnosis of NASH. Approximately 264 patients were to be randomized into 1 of 3 parallel treatment groups (300 mg NST-4016, 600 mg NST-4016 or Placebo); randomization was stratified by baseline biopsy fibrosis score (2 levels: F1 or F2 /F3).

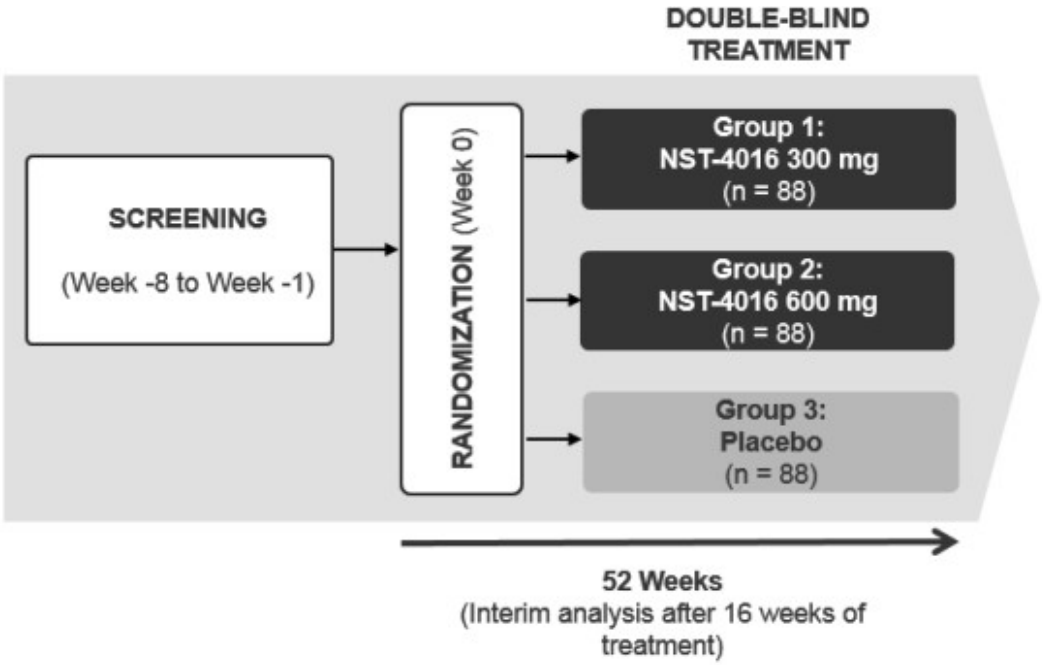
The study included a screening period, double-blind treatment period, and post-treatment follow-up. The Screening Visit could occur up to 8 weeks prior to randomization at Visit 2 (Week 0). The double-blind treatment period entailed 52 weeks of dosing with study drug and was to be followed by a Safety Follow-up Visit 2 weeks after the last dose of study drug. The study design is depicted in Figure 1 – Study Design.

Study drug dosing was to begin at Visit 2 (Week 0) and continue daily for 52 weeks. All patients were instructed to take 2 capsules of study drug (either 300 mg NST-4016 and/or matching placebo) orally in the morning with water once daily for 52 weeks. In the event that the Visit 9 biopsy was delayed, patients continued to administer study drug daily until the biopsy occurred, for up to 30 days after the originally projected Visit 9 date.

During the 52-week treatment period, patients were to return to the clinic for study visits every 1 to 3 months.

An external, independent Data Safety Monitoring Committee (DSMC) reviewed all available unblinded preliminary safety and biomarker results per the DSMC Charter.

Figure 1 – Study Design



In addition, the following data was reviewed during an interim analysis, which was performed after 90 patients completed 16 weeks of treatment: alanine aminotransferase (ALT), aspartate aminotransferase (AST) and AST/ALT ratio, Homeostatic Model Assessment for Insulin (HOMA-IR), Pro-C3, Enhanced Liver Fibrosis (ELF) panel, magnetic resonance imaging-proton density-fat fraction (MRI-PDFF), Iron-corrected T1 (cT1) and lipid parameters. The DSMC reviewed summary statistics and made recommendations which were shared blindly with the sponsor and clinical sites.

2.2.2 Sample Size Determination

A sample size of 88 patients per treatment group provides 80% power to detect a treatment effect of 22% (40% responder rate for active versus 18% placebo), assuming a dropout rate of 25%.

2.3 Study Endpoints/Parameters

2.3.1 Primary Efficacy Endpoint

The primary efficacy endpoint is the percentage of patients with resolution of NASH, defined as disappearance of ballooning (score = 0) with lobular inflammation score 0 or 1, and with no worsening of fibrosis.

2.3.2 Key Secondary Efficacy Endpoints

The key secondary efficacy endpoints are:

Percentage of patients with

- fibrosis improvement, defined as greater than or equal to one stage of fibrosis improvement and no worsening of steatohepatitis (inflammation/ballooning)
- resolution of NASH (defined as disappearance of ballooning (score = 0) with lobular inflammation score 0 or 1 AND fibrosis improvement (defined as greater than or equal to one stage of fibrosis improvement)

2.3.3 Other Secondary Efficacy Endpoints

Percentage of patients with:

- fibrosis improvement, defined as greater than or equal to one stage of fibrosis improvement
- 2-stage fibrosis improvement, defined as greater than or equal to two stages of fibrosis improvement

Change from baseline in:

- Liver enzymes - AST, ALT, alkaline phosphatase (ALP), gamma glutamyl transferase (GGT), and bilirubin
- Inflammation marker – high-sensitivity C-reactive protein (hsCRP)
- Fibrosis activity markers – N-terminal type 3 collagen propeptide (Pro-C3) and ELF panel
- HOMA-IR, a measure of insulin resistance and metabolic status.
- Composite Disease Activity Score – composite NASH score of inflammation, ballooning, fibrosis
- Nonalcoholic Fatty Liver Disease Activity Score (NAS)
- Individual histological scores for steatosis, ballooning, inflammation, and fibrosis
- Imaging parameters - MRI-PDFF

2.3.4 Exploratory Parameters

Exploratory biomarkers and endpoints of serum and tissue lipidomics and AI Histology from Path AI and/or HistoIndex and cT1 will be reported in a supplemental CSR if and when available.

2.3.5 Pharmacokinetic Parameters

The following PK parameters will be summarized:

- NST-4016 concentrations (all patients)

In addition, for the Intensive PK subset:

- AUC_{0-t}
- AUC_{0-inf}
- CL/F
- C_{max}
- T_{max}
- V_z/F

2.3.6 Safety Parameters

The safety endpoints include:

- Adverse events (AEs)
- Serious AEs (SAEs)
- Adverse events of special interest (AESI) – suspected drug induced liver injury.
- Vital signs
- Physical examinations
- Electrocardiograms (ECGs)
- Clinical laboratory parameters

3 STATISTICAL METHODOLOGY

3.1 General Considerations

3.1.1 Analysis Day

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

3.1.2 Analysis Visits

Scheduled visits will be assigned to analysis visits as recorded on the electronic case report form (eCRF). Unscheduled and early termination visits will be assigned to analysis visits according to the visit windows described in the following table:

Analysis Visit	Target Analysis Day	Low Analysis Day	High Analysis Day
Visit 2 (Week 0)	1	NA	NA
Visit 3 (Week 4)	28	2	49
Visit 4 (Week 10)	70	50	91
Visit 5 (Week 16)	112	92	140
Visit 6 (Week 24)	168	141	196
Visit 7 (Week 32)	224	197	252
Visit 8 (Week 40)	280	253	322
Visit 9 (Week 52)	364	323	371
Visit 10 (Week 54) Safety FUP	378	372	

If there is more than one assessment within a visit window, then the analysis visit will be assigned by the following priorities:

1. Use the assessment from the visit with the matching visit label.
2. Use the visit closest to the target visit day; in the case of ties, use the earlier visit.

Further details of analysis visit assignment will be provided in the Analysis Data Model (ADaM) Specifications

3.1.3 Definition of Baseline

Baseline is defined as the value at the Screening Visit for the liver biopsy and imaging data and as the predose value at Visit 2 (Week 0) for the biomarker data. For all other efficacy and safety endpoints, baseline is defined as the predose value at Visit 2 (Week 0); if this value is missing, then the latest predose value is used.

3.1.4 Summary Statistics

Categorical data will generally be summarized with counts and percentages of patients. 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, standard deviation, median, Q1, Q3, minimum, and maximum.

3.1.5 Hypothesis Testing

Given the primary and key secondary endpoints along with 2 active doses, hypothesis testing will be performed in a hierarchical order to preserve the overall Type 1 error.

The 600mg dose will be tested first vs placebo. The primary and the 2 key secondary endpoints will be tested in the order as presented in section 2.3.2. Testing of these 3 endpoints will next move to the 300mg dose arm vs placebo. If during the sequential testing of these 6 hypotheses, a p-value is non-significant (at alpha = 0.05), the remaining (from that point onwards) p-values will be considered nominal and significance cannot be claimed.

No multiplicity adjustments will be applied for other secondary endpoints and exploratory parameters analysis. Since the testing of the other secondary endpoints will not involve any control of Type 1 error, all associated p-values are deemed to be nominal.

3.1.6 Handling of Dropouts and Missing Data

Date Values

In cases of incomplete dates (e.g., AE, concomitant medication, and medical history start and/or stop dates), the missing component(s) will be assumed as the most conservative value possible. For example, if the start date is incomplete, the first day of the month will be imputed for the missing day and January will be imputed for the missing month. If a stop date is incomplete, the last day of the month will be imputed for the missing day and December will be imputed for the missing month. Incomplete start and stop dates will be listed as collected without imputation.

Date imputation will only be used for computational purposes such as treatment-emergent status. Actual date values, as they appear in the original case report forms (CRFs), will be presented within the data listings.

Non-Date Values

Missing values for the primary endpoint, Week 52 (or at the early termination (ET) visit, if applicable) liver biopsy will be imputed as nonresponders.

All other efficacy and safety data will be used according to availability, with no imputation for missing data.

3.2 Analysis Populations

3.2.1 Randomized Population

The Randomized population will include all patients who were randomized to the study.

3.2.2 Intent-to-Treat (ITT) Population

The ITT Population will include all patients who were randomized and received at least 1 dose of study drug.

3.2.3 Modified Intent-to-Treat (mITT) Population

The mITT Population will include all patients from the ITT Population who have valid baseline and Week 52 (or ET, if applicable) liver biopsy measurements; a valid Week 52/ET liver biopsy is

defined as a liver biopsy that has occurred within 90 days of last dose/Week 52 visit. This will be the primary population for efficacy analysis.

3.2.4 Per-Protocol (PP) Population

The PP Population will include all patients from the mITT Population who complete the 52-week treatment period without any protocol violations that could impact on the primary efficacy endpoint. Patients who have a valid week 52/EOT biopsy within 30 days of last dose will be included in the PP population. The PP Population, along with the reason for exclusion, will be finalized prior to study unblinding.

3.2.5 Trough Pharmacokinetics (PK) Population

The Trough PK Population will include all patients who received at least 1 dose of NST-4016 and have at least 1 measurable trough PK concentration.

3.2.6 The Intensive Pharmacokinetics (PK) Population

The Intensive PK Population will include all patients who received at least 1 dose of NST-4016 and have at least 1 measurable non-trough intensive PK sample.

3.2.7 Safety Population

The Safety Population will include all patients who received at least 1 dose of study drug.

3.3 Patient Data and Study Conduct

3.3.1 Patient Disposition

Counts and percentages of patients who were screened (signed informed consent), screen failed, and randomized will be summarized overall based on all screened patients. Reasons for screen failure will also be summarized.

Counts and percentages of patients who were randomized, withdrew early from the study, including primary reason for withdrawal, and those that completed the study will be summarized by treatment group and overall based on the Randomized Population.

For each scheduled visit, counts and percentages of patients who do not complete the visit, partially complete the visit in-person, or complete the visit virtually will be summarized by treatment group and overall. The denominator for calculating percentages will be based on the ITT Population.

Data listings for patient disposition and exclusion and inclusion criteria violations will be provided.

3.3.2 Protocol Deviations

Protocol deviations will be identified based on the clinical data as defined in the Protocol Deviation Plan. The Protocol Deviation Plan will define all protocol deviations as either CSR reportable or non-CSR reportable. Counts and percentages of patients with CSR reportable protocol deviations by deviation category will be summarized by treatment group and overall, for the Randomized Population. A similar summary will be provided for Coronavirus Disease 2019 (COVID-19) related CSR reportable protocol deviations. A listing of CSR reportable deviations will also be generated.

3.3.3 Analysis Populations

Counts and percentages of patients in each analysis population will be summarized by treatment group and overall based on the Randomized Population. Reasons for exclusion from PP Population will also be summarized.

A listing of patients included in each of the study populations will be provided.

3.3.4 Demographic and Baseline Characteristics

Demographic and baseline characteristics will be summarized with descriptive statistics or counts and percentage of patients as appropriate by treatment group and overall, for the ITT population. If they differ from the ITT Population, summaries will also be provided for the mITT Population, the and the PP Population.

Demographic characteristics include age, gender, ethnicity and race.

Baseline characteristics include:

- Body Mass Index (BMI)
- Alcohol use (current/former/never)

Baseline histology includes:

- Biopsy fibrosis score (F1, F2 or F3)
- Biopsy fibrosis score for stratification (2 levels: F1 or F2/F3)
- NAS Score and individual scores for steatosis, ballooning, inflammation and fibrosis
- Composite Disease Activity Score – composite NASH score of inflammation, ballooning and fibrosis.

Baseline chemistry includes:

- Liver enzymes (AST, ALT, ALP, GGT and bilirubin [Total, Direct and Indirect])
- Imaging parameters (MRI-PDFF and cT1)
- Biomarkers (HOMA-IR, hsCRP, Pro-C3 [COBAS] and ELF test)

All demographic, baseline characteristic, histology and chemistry data will be provided in by patient listings.

3.3.5 Medical History

Medical history will be coded to system organ class and preferred term using the Medical Dictionary for Regulatory Activities (MedDRA) version 22.0. Counts and percentages of patients with medical history by system organ class and preferred term will be summarized by treatment group and overall for the Randomized Population.

A listing of all medical history data will be provided.

3.3.6 Prior and Concomitant Medications

Prior and concomitant medications will be coded to World Health Organization (WHO) Drug Dictionary Global B3, March 2019. Prior medications include any medication taken within 6 months prior to screening. Concomitant medications include any medication taken on or after the date the patient signs informed consent.

Counts and percentages of patients taking prior and concomitant medications by Anatomical Therapeutic Chemical (ATC) class and preferred term will be summarized by treatment group and overall for the Safety Population.

All medications will be listed and flagged as either prior or concomitant (or both).

3.3.7 Study Drug Exposure and Compliance

Patients' exposure to study drug will be summarized with descriptive statistics for the Safety Population. 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.

Note that the exposure duration calculation is intended to describe the length of time a patient was exposed to study drug and therefore does not take study drug interruptions into account.

Exposure to study drug will be summarized by treatment group for the Safety Population using descriptive statistics. Furthermore, exposure will be summarized in the following categories:

- ≤4 weeks (≤28 days)
- >4 to ≤10 weeks (29 - 70 days)
- >10 to ≤16 weeks (71 - 112 days)
- >16 to ≤ 24 weeks (113 - 168 days)
- >24 to ≤ 32 weeks (169 – 224 days)
- >32 to ≤ 40 weeks (225 -280 days)
- >40 to ≤ 52 weeks (281 – 364 days)
- > 52 weeks

Summary statistics will be presented for overall compliance to study drug by treatment group. Percent compliance will be calculated as:

$$\frac{number\ of\ actual\ capsules\ taken}{number\ of\ expected\ capsules\ taken} \times 100$$

The number of actual capsules taken will be calculated as number of capsules dispensed – number of capsules returned. If study drug is not returned, the number of capsules returned will be considered 0 for the compliance calculation. The number of expected capsules will be derived from the dates of randomization and the earliest of the Visit 9 (Week 52) and ET dates, assuming 2 capsules are taken each day. Descriptive statistics will be used to summarize the compliance to the study drug by treatment group for the Safety Population. Additionally, the number and

percentages of patients in each treatment group will be provided with compliance in the following categories:

- <75%
- 75-125%
- >125%

Listings of drug accountability and derived exposure and compliance will be provided for the Safety Analysis Set.

3.4 Efficacy Assessment

All efficacy analyses will be performed using the mITT Population and repeated in for the ITT Population. Analyses of the primary and all secondary endpoints will also be repeated for PP Population. All efficacy data will be listed.

3.4.1 Primary Efficacy Endpoint

Primary Analysis

The primary efficacy endpoint is the percentage of patients with resolution of NASH without worsening of fibrosis. Patients without a valid Week 52 (or ET, if applicable) liver biopsy will be imputed as non-responders. The number and percentage of patients meeting the criteria for the primary endpoint will be summarized by treatment group.

The analysis of the primary endpoint will be performed using the data from the 600 mg NST-4016 and placebo groups using the Cochran-Mantel-Haenszel test, stratified by baseline biopsy fibrosis score (2 levels: F1 or F2/F3). The Cochran-Mantel-Haenszel test p-values will be reported, along with Odds Ratio and 95% confidence interval. Sample SAS code is below.

```
/************************************************************
TREATMENT: 0=Placebo, 1=NST-4016 600 mg
STATIFICATION_VARIABLE: liver fibrosis score(F1 versus F2 or F3)
OUTCOME: 1-resolution of NASH without worsening fibrosis; 0 - resolution of
NASH with worsening fibrosis or non-responders
COUNT: percentage of patients
************************************************************/

proc freq;
     table STRATIFICATION_VARIABLE*TREATMENT*OUTCOME / cmh;
     weight COUNT;
run;
```

Secondary Analyses

The key secondary analysis of the primary endpoint will be conducted using data for the 300 mg NST-4016 and placebo groups, again using the Cochran-Mantel-Haenszel test, stratified by baseline biopsy fibrosis score (2 levels: F1 or F2/F3).

Sensitivity Analyses

A sensitivity analysis for the primary endpoint will be carried out using a logistic regression model with treatment group (300 mg NST-4016, 600mg NST-4016, or placebo) and baseline biopsy fibrosis score (2 levels: F1 or F2/F3) as factors. Pairwise treatment comparisons will be estimated in this model as odds ratios (NST-4016 versus placebo) along with their 95% confidence intervals and p-values. Sample SAS code is below.

```
/*************************************************************************
TREATMENT: 0=Placebo, 1=NST-4016 600 mg, 2=NST-4016 300mg
STRATIFICATION_VARIABLE: liver fibrosis score (F1 versus F2 or F3)
OUTCOME: 1-resolution of NASH without worsening fibrosis;
         0 - resolution of NASH with worsening fibrosis or non-responders
*************************************************************************/

ods output ParameterEstimates=pa
           OddsRatios=or;

proc logistic;
     class TREATMENT (ref='0') STRATIFICATION_VARIABLE/ param=ref;
     model OUTCOME (event='1') = TREATMENT STRATIFICATION_VARIABLE;
run;
```

Additional sensitivity analyses of the analyses described above, CMH and logistic regression, will be conducted for the primary endpoint replacing the stratification factor of baseline biopsy score with the baseline ballooning score (1 or 2).

3.4.2 Secondary Efficacy Endpoints

Continuous efficacy variables will be summarized using descriptive statistics for the observed data and change and/or percent change from baseline by treatment group and visit as appropriate.

3.4.2.1 Responder analysis

The key secondary efficacy endpoints will be summarized and analyzed as described above for the primary efficacy endpoint.

The analysis of key secondary endpoints will also be tested in a hierarchical order as described in Section 3.1.5.

3.4.2.2 Nonalcoholic fatty liver disease activity score (NAS)

The NAS measures disease changes by scoring features of NASH during therapeutic trials. The total NAS represents the sum of scores for steatosis, lobular inflammation, and ballooning, and ranges from 0 to 8.

The change from baseline to Week 52/ET in NAS and individual histological scores (steatosis, ballooning, inflammation, and fibrosis) and a composite Disease Activity Score of inflammation, ballooning and fibrosis will be analyzed using an analysis of covariance model (ANCOVA). The model will be adjusted for treatment (300 mg NST-4016, 600 mg NST-4016, or placebo), biopsy

fibrosis score (2 levels: F1 or F2/F3), and respective baseline score. Pairwise treatment comparisons between each NST-4016 dose and placebo will be estimated using least square (LS) means, standard errors, 95% confidence intervals, and p-values. Sample SAS code is below.

```
/*****************************************************************
TREATMENT: 0=Placebo, 1=NST-4016 600 mg, 2=NST-4016 300mg
STRATIFICATION_VARIABLE: liver fibrosis score (F1 versus F2 or F3)
BASE: NAS baseline value
CHANGE: change from baseline to Week 52/ET in NAS
*****************************************************************/


proc glm data=mydata;
     class TREATMENT STRATIFICATION_VARIABLE;
     model CHANGE= TREATMENT STRATIFICATION_VARIABLE BASE;
     lsmeans TREATMENT/pdiff cl;
run;
```

A sensitivity analysis will also be conducted for the NAS and individual histological scores withi an ANCOVA model replacing the stratification factor of baseline biopsy score with the baseline ballooning score (1 or 2).

3.4.2.3 Liver Enzymes

Liver enzymes are assessed at Screening, Visit 2 (Week 0), Visit 3 (Week 4), Visit 4 (Week 10), Visit 5 (Week 16), Visit 6 (Week 24), Visit 7 (Week 32), Visit 8 (Week 40), Visit 9 (Week 52), Visit 10 (Week 54) and at the ET Visit.

The change and percent change from baseline in AST, ALT, ALP, GGT and total bilirubin at each scheduled visit will be analyzed using a Mixed Model Repeated Measures (MMRM) model, with change (or percent change) from baseline as the dependent variable, adjusting for treatment (300 mg NST-4016, 600 mg NST-4016, or placebo), visit, treatment-by-visit interaction, biopsy fibrosis score (2 levels: F1 or F2/F3); and baseline score. The Restricted Maximum Likelihood (REML) estimation approach will be used with an unstructured covariance matrix. The LS means, standard errors, 95% confidence intervals, and p-values for the pairwise treatment comparisons between each NST-4016 dose and placebo will be estimated for each visit.

At visits where the fasted status of the patient is recorded (Visit 2 (Week 0), Visit 5 (Week 16) and Visit 9 (Week 52)), only samples taken in a fasted state will be included. All results from other visits will be included. An additional analysis will be carried out to include patients who we fasted at all three visits where fasted status is recorded.

3.4.2.4 Inflammation Marker: hsCRP

High sensitivity C-reactive protein (hsCRP) is measured at Visit 2 (Week 0), Visit 5 (Week 16) and Visit 9 (Week 52)/ET.

To be included in the analysis, the patient must have been in a fasted state at the time of the assessment.

The change and percent change from baseline in hsCRP and the log-transformation will be analyzed using a similar MMRM model as described for the Liver Enzymes.

3.4.2.5 Fibrosis activity markers: Pro-C3 and ELF Test

Pro-C3 (COBAS) and ELF test are measured at Visit 2 (Week 0), Visit 5 (Week 16) and Visit 9 (Week 52)/ET.

To be included in the analysis, the patient must have been in a fasted state at the time of the assessment.

The change and percent change from baseline in Pro-C3 (COBAS) and ELF will be analyzed using a similar MMRM model as described for the Liver Enzymes.

3.4.2.6 Measure of insulin resistance and metabolic status: HOMA-IR

HOMA-IR is measured at Visit 2 (Week 0), Visit 5 (Week 16) and Visit 9 (Week 52)/ET.

To be included in the analysis, the patient must have been in a fasted state at the time of the assessment.

The change and percent change from baseline in HOMA-IR will be analyzed using a similar MMRM model as described for the Liver Enzymes.

3.4.2.7 MRI-PDFF

Hepatic imaging by MRI will be performed at Screening, Visit 5 (Week 16) and Visit 9 (Week 52)/ET. Images will be assessed centrally for MRI-PDFF.

The change from the baseline to Week 16 and Week 52/ET in MRI-PDFF will be performed using similar ANCOVA models as described above. The change from baseline to Week 16 and Week 52/ET will be analyzed in separate models.

3.4.3 Exploratory Biomarker Analysis

As part of the exploratory analysis, cT1 will be analyzed as described for MRI-PDFF above.

Lipodomics data will be available for a sub-set of the patients; data will be analyzed using similar ANCOVA or MMRM models as described for the secondary endponts.

The AI Histology data will be compared with the liver biopsy results from the manual reads.

3.5 Pharmacokinetic Assessment

3.5.1 Sample Collections for Pharmacokinetic Analysis

For the Trough PK Population, PK blood samples of NST-4016 will be collected pre-dose at Visit 2 through to Visit 9 (or the ET Visit, if applicable).

For the Intensive PK Population, PK blood samples of NST-4016 will be collected at the following timepoints, in relation to the first dose of study drug on Visit 2: pre-dose (within 1 hour prior to first

dose), followed by 1, 2, 3, 4, 6, and 8 hours (+/- 10 minutes) post-dose (Day 0) and at 24hours post-dose (+/- 1 hour, but prior to the second dose of study drug on Day 1).

3.5.2 Handling Missing or Below the Lower Limit of Quantification Data

For PK concentration data, if the actual sampling time is missing, but a valid concentration value has been measured, the concentration value will be flagged, and the scheduled time point may be used for the calculation of PK parameters.

In cases of missing pre-dose concentrations for Visit 2 the missing components may be assumed as zero. For all other cases, the missing data will not be imputed.

At Visit 2, PK blood concentration values that are below the lower limit of quantification will be presented as “BLQ” in the concentration data listing.

For the individual concentration and PK parameter calculation, the following rules will be applied:

- If one or more BLQ values occur before the first measurable concentration, they will be assigned a value of zero.
- If BLQ values occur between measurable concentrations in a profile, the BLQ should be omitted (set to missing).
- If BLQ values occur after the last measurable concentration in a profile, the BLQ should be omitted (set to missing).
- For the concentration summary and mean plot preparation, the following rules will be applied: Mean concentration at any individual time point will only be calculated if at least half of the subjects have valid values (i.e. quantifiable and not missing) at this time point
- In cases where a mean value is not calculated due to the above criterion not being met, the value will be set to missing
- BLQ values will be set to zero

3.5.3 Pharmacokinetic Concentration

For the Trough PK Population, trough plasma PK concentrations, defined as the plasma PK concentration taken prior to the administration of study drug, at Visits 3 to 9 will be listed and summarized descriptively by visit and treatment. Mean (+/- SD) trough plasma PK concentrations will be plotted by visit and treatment group on a linear and semi-logarithmic scale against nominal sampling time points. Individual trough plasma PK concentrations will be plotted on linear and semi-logarithmic scales against actual sampling time points relative to dosing time.

For the Intensive PK Population, Individual plasma PK concentrations will be listed and summarized by treatment group at each nominal time point descriptively. Mean (+/- SD) plasma PK concentrations will be plotted by treatment group on linear and semi-logarithmic scales against nominal time. Individual plasma PK concentrations will be plotted on linear and semi-logarithmic scales by treatment group against actual sampling time points relative to dosing time.

For all plots, LLOQ will be plotted as a reference line.

3.5.4 Pharmacokinetic Parameters

For the Intensive PK Population at Visit 2, the following plasma PK parameters of NST-4016 will be determined using non-compartmental methods as appropriate:

Parameter	Description
C_{max}	Maximum observed plasma concentration; determined directly from the concentration time profile; if the maximum plasma concentration occurs at more than one time point, C_{max} is defined as the first maximum value
T_{max}	Time to C_{max}; If the maximum value occurs at more than one time point, T_{max} is defined as the first time point with this value.
λz	Apparent first-order terminal elimination rate constant calculated from a semi-log plot of the plasma concentration versus time curve. The parameter will be calculated by linear least-squares regression analysis using points in the terminal log-linear phase.
$t_{1/2}$	Apparent first-order terminal elimination half-life; calculated as ln(2)/λz
AUC_{0-t}	Area under the plasma concentration-time curve from time 0 to the time of the last observed concentration
AUC_{0-inf}	Area under the plasma concentration-time curve from time 0 extrapolated to infinity; calculated as (AUC0-last+ Clast/λz)
$AUC_{\%extrap}$	Percent of AUC0-infextrapolated, represented as (1 -AUC0-t/AUC0-inf)*100
CL/F	Apparent total plasma clearance after oral administration; calculated as Dose/AUC0-inf
Vz/F	Apparent volume of distribution during terminal elimination phase after oral administration; calculated as Dose/[λz*AUC0-inf]

Actual collection times will be used for the calculation of PK parameters. AUCs will be calculated using linear up log down method.

λz will not be presented for subjects who do not exhibit a terminal elimination phase in their concentration-time profiles. In order to estimate λz, linear regression of concentration in logarithm scale versus time will be performed using at least 3 data points. Uniform weighting will be selected to perform the regression analysis.

The constant λz will not be assigned if one of the following occurs:

1. The terminal elimination phase is not linear (as it appears on a semi-logarithmic scale).
2. The terminal elimination rate constant indicates a positive slope (λz > 0).
3. Tmax is one of the three last data points.

The constant λz and its derived parameters will be listed but excluded from statistical analysis if one of the following occurs:

1. The adjusted regression coefficient (R^2) is less than 0.8.
2. The AUC%extrap exceeds 20%.

If λz is not assigned, the values of any associated PK parameters (e.g. $AUC_{0\text{-inf}}$, $AUC_{\%extrap}$, $t_{1/2}$, CL/F, or V_z/F) will not be calculated.

No PK parameters will be calculated for subjects with 2 or fewer detectable concentrations in their PK profile.

Plasma PK parameters will be summarized by treatment using descriptive statistics for the Intensive PK Population.

3.6 Safety Assessment

Safety data include AEs, vital signs, physical examinations, ECGs and clinical laboratory parameters and will be summarized by actual treatment received based on the Safety Population.

3.6.1 Adverse Events (AEs)

An adverse event is defined as any untoward medical occurrence in a clinical investigation patient administered a pharmaceutical product, which does not necessarily have a causal relationship with this treatment. An adverse event can therefore be any unfavorable and/or unintended sign (including an abnormal laboratory finding), symptom, or disease temporally associated with the use of an investigational medicinal product, whether or not related to the investigational medicinal product. A treatment emergent AE (TEAE) is defined as a new or worsening adverse event after the first dose of study drug.

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

An overview of AEs will be provided including counts and percentages of patients and number of events by treatment group and overall for Safety Population (and event counts) with the following:

- Any Adverse Event
- Any TEAEs
- Maximum severity of TEAEs
- Study drug related TEAEs
- Maximum severity of study drug related TEAEs
- TEAEs leading to discontinuation of study drug
- Any TEAEs of special interest (TEAESI)
- Any serious TEAEs (TESAEs)
- Study drug related TESAEs
- TESAE/TEAESIs leading to discontinuation of study

- TESAEs leading to death
- TESAEs related to COVID leading to death

The TEAEs described above will be summarized separately by System Organ Class and Preferred Term (and by maximum severity).

Listings will be presented specifically for all TEAEs, TEAESI, TESAEs, TEAEs Leading to Discontinuation of Study Drug, TEAEs Leading to Discontinuation of Study and TESAEs Leading to Death.

3.6.2 Clinical Laboratory Tests

Fasting Hematology, biochemistry, coagulation and lipid panels and urinalysis will be collected at all study visits. A list of laboratory tests to be performed is included in Protocol Appendix B.

Descriptive statistics of each clinical laboratory parameter will be presented for baseline values and for values and the change and percent change from the baseline at each post-baseline visit by treatment group and overall.

Values outside the normal range will be categorized as High (i.e., above the normal range) or Low (i.e., below the normal range) based on the central laboratory's normal reference range. The number and percentage of patients with values outside the normal range will be summarized by parameter for baseline and for each post-baseline visit.

Shift tables will also be presented for baseline versus maximum-baseline values and baseline versus minimum post-baseline values.

All safety laboratory data will be presented in patient listings.

3.6.3 Vital Signs

Vital signs will be measured at all study visits. Vital signs parameters include blood pressure, heart rate and body temperature. Three replicate blood pressure readings will be taken at each visit and the average of these measurements will be used in the analysis.

Vital signs parameters will be summarized by visit, treatment group and overall. Change and percent change from baseline at each post-baseline visit will be presented for each parameter.

All vital sign data will be presented in patient listings.

3.6.4 Anthropometric measures

Height is measured at the Screening Visit only. Weight will be measured at each study visit. BMI will be derived for each study visit using the weight measured at that visit and the height measured at the screening visit. Anthropometric data will be summarized by visit, treatment and overall. Change from baseline at each post-baseline visit will be presented for weight and BMI.

All anthropometric data will be presented in patient listings.

3.6.5 Electrocardiograms

Single 12-lead ECGs will be performed at Screening, Visit 2 (Week 0), Visit 5 (Week 16), Visit 7 (Week 32), Visit 9 (Week 52), and at the Safety Follow-up or ET Visit. ECG parameters include Heart Rate, PR Interval, QRS Duration, QT Interval, QTcF Interval and RR Interval.

ECG parameters will be summarized by visit, treatment group and overall. Change and percent change from baseline at each post-baseline visit will be presented for each parameter.

Additionally, the number and percentages of patients in each treatment group will be provided with QTcF in the following categories:

- Absolute values:
    - QTcF Interval > 450 msec
    - QTcF Interval > 480 msec
    - QTcF Interval > 500 msec
- Change from baseline:
    - QTcF increase from baseline > 30 msec
    - QTcF increase from baseline > 60 msec

Overall interpretation will also be summarized by visit.

All ECG data will be presented in patient listings.

3.6.6 Physical Examinations

Full physical examinations (including general appearance; eyes, ears, nose, and throat; head and neck; chest and lungs; cardiovascular; abdomen; musculoskeletal; lymphatic; dermatological; neurological systems; mental status; and extremities) will be performed at Screening; Visit 2 (Week 0), Visit 9 (Week 52) and at the Safety Follow-up or ET Visit. Abbreviated physical examinations will be performed at all other study visits.

All physical examination data will be listed.

4 DATA SAFETY MONITORING COMMITTEE

An external, independent Data Safety Monitoring Committee (DSMC) was to review all available unblinded preliminary safety and biomarker results, as described in the DSMC Charter. The DSMC Charter describes the overall guidelines, composition and roles, and responsibilities of the independent DSMC, timing of meetings, methods of providing information to and from the DSMC, frequency and format of meetings, data analysis recommendations, and DSMC relationships with other parties participating in the conduct of this study. Patient accrual continued throughout the period of the DSMC review. The DSMC was to advise on the accrual of the remaining patients per protocol or any amendments that are necessary for safety reasons.

Patients, investigators, site staff and in general all personnel directly involved in the conduct of the study were to remain blinded to the patients' treatment assignment until the completion of the study.

5 ANALYSIS TIMING

5.1 Interim Analysis

An interim analysis was planned for this study after 90 patients had completed 16 weeks of treatment. This analysis was reviewed by the DSMC according to the DSMC Charter. The conclusions and recommendations of the DSMC were shared blindly with the Sponsor and clinical sites. A separate interim analysis plan was finalized prior to conducting the interim analysis.

5.2 Pre-Final Analysis

After the database is locked and exclusions from analysis populations have been finalized, the randomized treatment assignments will be unblinded and the pre-final analysis will be generated. Pre-final tables, figures and listings (TFLs) will be provided approximately 3 weeks after database lock.

5.3 Final Analysis

After all comments on the pre-final analysis have been resolved and the study database is declared final, the final analysis will be generated. Final TFLs will be provided approximately 1 week after the study database is declared final. If there were no changes to the pre-final analysis or the study database, the pre-final TFLs may be considered final. In addition to TFLs, SDTM data and ADaM data along with associated files will be provided. Associated files may include: annotated CRFs, Standard Data Tabulation Model (SDTM) specifications, SDTM programs, ADaM specifications, ADaM programs, TFL programs, and Clinical Data Interchange Standards Consortium (CDISC) Define packages for both SDTM and ADaM data.

6 CHANGES FROM PROTOCOL-SPECIFIED STATISTICAL ANALYSES

The planned changes to the protocol-specified statistical analyses, as detailed in this SAP are:

- Key secondary efficacy endpoints added for different definitions of response – fibrosis improvement and resolution of NASH with fibrosis improvement.
- cT1 demoted to exploratory endpoint (from secondary) as there is uncertainty over clinical significance.
- The primary population for all efficacy analysis updated to mITT Population, with all analyses to be repeated for the ITT Population.
- No imputation to be performed for the continuous secondary endpoints.
- ALP added to list of liver enzymes to be analyzed as secondary endpoints.

7 PROGRAMMING SPECIFICATIONS

Analyses will be performed using SAS® version 9.4 or higher. All available data will be presented in patient data listings which will be sorted by patient and visit date as applicable.

Phoenix WinNonlin version 8.3 or higher will be used to calculate PK parameter estimates. PK parameter estimates will also be calculated via SAS and verified with the Phoenix WinNonlin results.

Detailed Programming Specifications will be provided in a separate document(s).