NOVARTIS

Clinical Development

LJN452

CLJN452A2202 / NCT02855164

A randomized, double-blind, placebo controlled, 2- part, adaptive design, multicenter study to assess safety, tolerability and efficacy of tropifexor (LJN452) in patients with non-alcoholic steatohepatitis (NASH)

FLIGHT-FXR

Statistical Analysis Plan (SAP)

Author:	Statistician, [redacted]

Document type:	SAP Documentation
Document status:	Final 4.0
Release date:	02-July-2019
Number of pages:	46

Document History – Changes compared to previous final version of SAP

Time point	Reason for update	Outcome for update	Section and title impacted (Current)
Protocol Amendment 2	Clarification for R programming	Additional specification added to allow for more flexible programming approach for MCP-Mod implementation	2.5.2 Statistical hypothesis, model, and method of analysis
Protocol Amendment 2	Clarification for programming	Added that posterior probability plot should be based on flat prior (using only likelihood of observed data)	2.5.2 Statistical hypothesis, model, and method of analysis
Protocol Amendment 2	Information from phase I study	Changed MMRM for FGF19 and C4 to simple ANCOVA at week 6 because elevation is expected to be seen only within few hours post-dose.	2.7.1 Secondary endpoints and 2.7.2 Statistical hypothesis, model, and method of analysis
Protocol Amendment 2	Better understanding of assessment schedule	Changed MMRM for Fibroscan, ELF and fibrosis biomarker test to simple ANCOVA because there is only one post-baseline assessment.	2.7.1 Secondary endpoints
Protocol Amendment 2	Clarification	Additional specification that log(ratio) from baseline is analyzed for lipids	2.7.1 Secondary endpoints
Protocol Amendment 2	Clarification	Deleted sentence that any efficacy assessment after dose reduction would be set to missing for respective analyses, because this was only meant for MMRM based analyses.	2.7.3 Handling of missing values/censoring/ discontinuations
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
Protocol Amendment 2	Clarification	Aligned bullet points in AE section with TFL shells	2.8.1 Adverse events
Protocol Amendment 2	New mandatory text	Added mandatory text for EudraCT disclosure.	2.8.1 Adverse events
Protocol Amendment 2	Update	AE coding / grading updated as per Amendment 1 and approved eCRS	5.2
Protocol Amendment 2	Protocol deviation	Modified SAF definition to account for dosing errors	2.2
Protocol Amendment 2	Protocol Amendment 2	Added reporting event “End of Part A”	1, 2.1, 2.14

Time point	Reason for update	Outcome for update	Section and title impacted (Current)
Protocol Amendment 2	Clarification	Additional specification for geographical region as a possible covariate, in case Japan/Non-Japan will be used as stratification factor in Part B	2.1
Protocol Amendment 2	Protocol Amendment 2	Moved analysis of relative change of liver fat to primary variables.	2.5.1, 2.5.2, 2.5.3, 2.7.1, 2.7.3, 3.1.2
Protocol Amendment 2	DMC request	Added individual lipid profile plots and summary statistics for FGF19 and C4 to 1st interim analysis / DMC review	2.14, TFL shells
[redacted]	[redacted]	[redacted]	[redacted]
Protocol Amendment 2	Clarification	Split regression model for anthropometric assessments into MMRM and ANCOVA because there are no repeated post-baseline assessments of WTH ratio	TFL shells, 2.7.1
Protocol Amendment 2	Request from GPMD, BSL	Added subgroup table of lipid summary by use of lipid reducing concomitant medication (and definition)	TFL shells, 5.3
[redacted]	[redacted]	[redacted]	[redacted]
Protocol Amendment 3	Protocol Amendment 3	All sections updated to include study Part C with additional reporting events, objectives, endpoints and doses	All sections, TFL shells
Protocol Amendment 3	Change of PD criteria definition	OTH11 was changed and OTH14 added. OTH11 no longer results in exclusion from any analyses, but instead OTH14 does.	5.5
Prior to Week 12 Part C DBL	Protocol Amendment 3	Title updated	Title page
Prior to Week 12 Part C DBL	Unexpected subject journey	Handling of data from subjects who discontinued study treatment but stayed in study with regular visit schedule	2.4.2, 2.8.1, 2.8.3
Prior to Week 12 Part C DBL	Clarification	Precise definition of biopsy based endpoints	5.3.11
Prior to Week 12 Part C DBL	DMC recommendation	More detailed analysis of Pruritus	2.8.1.1

Time point	Reason for update	Outcome for update	Section and title impacted (Current)
Prior to Week 12 Part C DBL	Leaning	Analysis of ALT and AST spikes removed for Part C	2.8.3
Prior to Week 12 Part C DBL	Leaning	Biomarker listing removed	2.12

Table of contents

List of abbreviations

AE	Adverse event
ALP	Alkaline phosphatase
ALT	Alanine transaminase
ANCOVA	Analysis of covariance
[redacted]	[redacted]
APTT	Activated partial thromboplastin time
AST	Aspartate transaminase
ATC	Anatomical Therapeutic Classification
AUC	Area Under the Curve
BMI	Body mass index
CI	Confidence interval
[redacted]	[redacted]
CSR	Clinical Study report
CTC	Common Toxicity Criteria
CTCAE	Common Terminology Criteria for Adverse Events
CV	Coefficient of variation
DMC	Data Monitoring Committee
eCRF	Electronic Case Report Form
eCRS	Electronic Case Retrieval Sheet
eGFR	Estimated glomerular filtration rate
ELF	Enhanced liver fibrosis panel
EOS	End of study
EOT	End of treatment
FAS	Full analysis Set
FIB-4	Fibrosis-4 score
GGT	Gamma glutamate transaminase
HDL	High density lipoprotein
[redacted]	[redacted]
INR	International Normalized Ratio
ITT	Intent-to-treat
LDL	Low density lipoprotein
LLOQ	Lower limit of quantification
MAR	Missing at random
MedDRA	Medical Dictionary for Drug Regulatory Affairs
MMRM	Mixed model repeated measures
MRI	Magnetic resonance imaging
NAFLD	Non-alcoholic fatty liver disease
NAS	NAFLD activity score
NASH	Non-alcoholic steatohepatitis
OCA	Obeticholic acid
PD	Pharmacodynamics
[redacted]	[redacted]

[redacted]	[redacted]
PK	Pharmacokinetics
PRO	Patient-reported Outcomes
PT	Preferred term / Prothrombin time
[redacted]	[redacted]
RAN	Randomized set
SAE	Serious adverse event
SAF	Safety set
SAP	Statistical Analysis Plan
SCR	Screened set
SD	Standard deviation
SMQ	Standardized MedDRA Query
SOC	System Organ Class
TBC	To be confirmed
TBL	Total bilirubin
TFLs	Tables, Figures, Listings
ULN	Upper limit of normal range
ULOQ	Upper Limit of Quantification
VAS	Visual analogue scale
WHO	World Health Organization

1 Introduction

The purpose of this document is to provide the detailed implementation of statistical analysis plan for study LJN452A2202. The analysis will result in one final study report when all patients have completed the study (Part A, Part B and Part C). In addition, this SAP provides details for the interim analysis (when at least 90% of the patients have completed the Week 8 assessment in Part A) and regular safety DMC analyses, as well as analyses at the end of Part A, at the end of Part B and when all patients have completed the Week 12 visit of Part C.

1.1 Study design

This is a randomized, double-blind, placebo-controlled, multicenter, parallel-group, dose finding, 3-part, adaptive, study to assess the safety, tolerability and efficacy of six doses of tropifexor as compared to placebo in patients with non-alcoholic steatohepatitis (NASH).

Patients can be included if they have either histologic evidence of NASH or phenotypic diagnosis of NASH (In Parts A & B), as further specified in Section 4.1 of the protocol. Histologic evidence of NASH (with fibrosis stage 2 or 3), per central read, during the Screening period or within 6 months of randomization is required for all patients in Part C. The composite criterion for phenotypic diagnosis has been shown to be nearly as precise as locally read liver biopsy which can be frequently re-read as 'non-NASH' [Neuschwander-Tetri et al (2015)].

The study will start with screening and enrolling patients for Part A. Screening will continue without pause, even after all patients for Part A have been enrolled, but randomization for Part B will not start until after the interim analysis in Part A. When ≥ 90% of the patients in Part A have completed 8 weeks of treatment, an interim analysis will be performed using all available data to allow for the DMC to recommend dose selection for the arms in Part B (see protocol Section 3.5 and Section 9.7). The treatment arms of Part A are planned to be completed without adaptation.

Randomization for Part B will only be started after the DMC recommendations on the dose(s) to be used in Part B are implemented by the Sponsor.

Part C was introduced as a result of the DMC recommendation to pursue doses > 90 µg after the planned interim analysis and DMC review of the Part A data, as well as the preclinical and pharmacokinetic data described in protocol section 3.3. Randomization into Part C will commence after Part B randomization is complete.

Patients in **Part A** will be assigned at the baseline visit to one of the following 5 treatment arms in a ratio of 1:1:1:1:1 in a blinded manner. Placebo capsules will be given in each treatment arm where necessary to maintain blinding:

Arm A: Once daily (morning, fasting) treatment with 10 µg tropifexor for 12 weeks

Arm B: Once daily (morning, fasting) treatment with 30 µg tropifexor for 12 weeks

Arm C: Once daily (morning, fasting) treatment with 60 µg tropifexor for 12 weeks

Arm D: Once daily (morning, fasting) treatment with 90 µg tropifexor for 12 weeks

Arm E: Once daily (morning, fasting) treatment with matching placebo capsules for 12 weeks

Patients in **Part B** will be assigned at the baseline visit to one of the following 3 treatment arms in a ratio of 15:4:5 in a blinded manner. The doses to be used in Part B were decided based on data from the first interim analysis and DMC consultation. Placebo capsules will be given in each treatment arm where necessary to maintain blinding.

Arm F: Once daily (morning, fasting) treatment with 90 µg tropifexor.

Arm G: Once daily (morning, fasting) treatment with 60 µg tropifexor.

Arm H: Once daily (morning, fasting) treatment with matching placebo for 12 weeks

Patients in **Part C** will be assigned at the baseline visit to one of the following 3 treatment arms in a ratio of 1:1:1 in a blinded manner. The doses used in Part C were determined based on preclinical toxicology data, pharmacokinetic modeling and DMC recommendations based on Part A data. Placebo capsules will be given to maintain blinding.

Arm I: Once daily (morning, fasting) treatment with 140 µg tropifexor for 48 weeks

Arm J: Once daily (morning, fasting) treatment with 200 µg tropifexor for 48 weeks

Arm K: Once daily (morning, fasting) treatment with matching placebo for 48 weeks

In order to maintain the blind, placebo capsules matching tropifexor 10, 30 and 100 µg capsules will be given to patients, so that all patients will receive 3 capsules per day.

Part A: Randomization will be stratified in Part A by BMI at baseline (<30 kg/m^2 or ≥ 30 kg/m^2 for patients with an Asian race or < 35 kg/m^2 or ≥ 35 kg/m^2 for all other patients). The race is based on the race the patient self-reports as captured on the demography eCRF.

Part B: Randomization will be stratified in Part B by BMI at baseline (<30 kg/m^2 or ≥ 30 kg/m^2 for patients with an Asian race or < 35 kg/m^2 or ≥ 35 kg/m^2 for all other patients). The race is based on the race the patient self-reports as captured on the demography eCRF. Randomization in Part B will also be stratified by Japanese or non-Japanese, Japanese defined as patients residing in Japan.

Part C: Randomization in Part C will be stratified by Fibrosis level (F2/F3) as determined in the Screening biopsy by the Central Reader and by Type 2 Diabetes status (yes/no) as reported on the Medical History eCRF. To ensure balance of patients in Japan among the 3 treatment arms, the following stratum levels will be implemented (Japanese subgroup is too small for additional stratification):

1. Japan
2. Non-Japan, F2, T2D=No
3. Non-Japan, F2, T2D=Yes
4. Non-Japan, F3, T2D=No
5. Non-Japan, F3, T2D=Yes

The primary analysis timepoint for efficacy assessments is Week 12 (end of treatment in Parts A and B) in all three study parts.

1.2 Study objectives and endpoints

Table 1-1 Objectives and related endpoints

Objective	Endpoint	Time Frame
Primary		
To determine safety and tolerability of different doses of LJN452	Occurrence of SAE, AE resulting in discontinuation of study treatment and/or dose reductions, and AE of special interest	up to End of Study [1]
To determine the dose relationship of LJN42 on markers of hepatic inflammation in NASH (ALT and AST)	Change from baseline to Week 12	up to Week 12
To determine the dose-response relationship of LJN452 on liver fat content by changes in quantitative MRI determined fat	Change from baseline to Week 12 in % of fat in the liver assessed using MRI	baseline, Week 12
Secondary		
To determine the effect of different doses of LJN452 on anthropometric assessments (weight, BMI, waist-to-hip (WTH) ratio) after 12 weeks of treatment	Changes from baseline to Week 12	up to Week 12
To determine the dose-response relationship of LJN452 on FGF19 over time, a marker of FXR target engagement in the gut, and C4, a marker of hepatic target engagement	Changes from baseline to Week 12	up to Week 12
To determine the dose-response relationship of LJN452 on markers of liver fibrosis commonly available such as Fibroscan®, enhanced liver fibrosis panel (ELF), and fibrosis biomarker test (originally known as Fibrotest®/ FibroSure®)	Changes from baseline to Week 12	up to Week 12
To determine the dose-response relationship of LJN452 on GGT, a marker of cholestasis	Changes from baseline to Week 12	up to Week 12
To determine the effect of LJN452 on fasting lipid profile	Changes from baseline to Week 12	up to Week 12
To determine the pharmacokinetics (PK) of LJN452	C_{trough}, C_{2h}	up to Week 12
To determine the effect of LJN452 compared to placebo with respect to occurrence of potential itch based on a visual analog scale (VAS) rating scale	The score (distance from left) on the VAS will be recorded by the patient marking with a line. The distance marked will be converted to a score between 0 and 10	up to Week 12

Objective	Endpoint	Time Frame
To determine effects of LJN452 on primary endpoints in the subset of patients who have historical biopsy data, both overall and by fibrosis score and/or NAS score subsets as feasible (based on the extent of available data)	Subgroup analysis with respect to: occurrence of SAE, AE resulting in discontinuation of study treatment and/or dose reductions, and AE of special interest, relative change from baseline to Week 12 in % of fat in the liver assessed using MRI, and change from baseline to Week 12 of ALT and AST and GGT	up to Week 12
Additional Secondary Endpoints for Part C: To demonstrate the efficacy of tropifexor in patients with NASH and F2/F3 fibrosis as assessed by histological improvement from baseline after 48 weeks of treatment compared to placebo	Proportion of patients who have at least a one point improvement in fibrosis without worsening of steatohepatitis at Week 48 compared to baseline Proportion of patients with resolution of steatohepatitis without worsening of fibrosis at Week 48 compared to baseline	up to Week 48
To determine the effect of tropifexor on markers of hepatic inflammation in NASH (ALT and AST) in Part C	Change from baseline to Week 48	up to Week 48
To determine the effect of tropifexor on liver fat content by changes in quantitative MRI determined fat in Part C	Relative change from baseline to Week 48 in % of fat in the liver assessed using MRI	baseline, Week 48
To determine safety and tolerability of different doses of tropifexor, adjusted for length of exposure	Exposure adjusted incidence of SAE, AE resulting in discontinuation of study treatment and/or dose reductions, AE of special interest and other AE	up to Week 52

[1] End of follow-up in Parts A and B can be Week 16 or the flexible Week 20 visit ; End of follow-up in Part C can be Week 52 or the flexible Week 56 visit

2 Statistical methods

2.1 Data analysis general information

This SAP includes analyses planned for the interim analyses as well as safety DMC reports.

Analyses will be run using SAS 9.3 or newer. Other statistical software and programming languages, such as R, may be used as well, in the newest version available in the validated environment.

Summary tables will be presented by treatment group and analysis visit (as applicable) using descriptive statistics. These include absolute and relative frequencies for categorical variables. Continuous variables will be summarized by arithmetic mean, standard deviation, minimum, maximum, median and first and third quartile per default (“statset1”). Where indicated, geometric mean and coefficient of variation will also be displayed, and the ratio instead of percentage change (“statset2”).

The final analysis will be performed after the final clinical database lock, when all patients have completed the study (Parts A, B and C), and will thefore include all collected data.

The final (end-of-study) analysis will be conducted on all patient data collected up to the Week 16/52 visit (EOS/Visit 299) or the premature treatment discontinuation visit (i.e., EOT (Week 12/48 visit, Visit 199) for patients discontinuing during the treatment period or EOS (Week 16/52 visit, Visit 299) for patients discontinuing during the Follow-up 1 period).

Data collected during the Follow-up 2 period will be described in a separate supplement to the primary CSR.

Unless noted otherwise, Part A and Part B data will be combined and analyzed jointly after end of Part B. Part C data will be analyzed separately except where indicated.

The first interim analysis, with the purpose to determine doses for Part B, will include all available clean data after at least 90% of the patients (67) have completed Week 8 in Part A or discontinued prior to Week 8. Of note, all clean data and complete data up to the date of the last patient with a Week 8 will be included as available. Data domains to be cleaned and required for the interim analysis will de described in the Data management Plan.

A second interim analysis (End-of-Part A analysis) will be performed when all subjects of Part A have completed the Week 16 visit, including only the subjects randomized to Part A. This analysis will not be performed by an independent team. However, as Part A and Part B cohorts use separate sections of the randomization list, only Part A treatment allocation information will be made available to the statisticians and programmers at this stage. The analyses performed will be a subset of the analyses planned for the final CSR and will be defined in a separate tracking sheet. A similar process will be applied after completion of Part B, when the joint analysis of Part A and B data will be performed.

At the Week 12 interim analysis of Part C, the Novartis study team and external statisticians and programmers working on the study will be unblinded to Part C treatment allocation, i.e. the Part C Week 12 analysis will not be performed by an independent team.

Data cutoffs for the safety DMC analyses will occur approximately every 6 months after start of randomization into Part B and Part C, as appropriate.

For the Part B and Part C analyses, geographical region will be included as a covariate in addition to the stratification factor (BMI category) in some of the analyses. Geographical regions are defined as follows, with countries known to participate in Part A and / or expected for Part B:

- Europe: Austria, France, Germany, Netherlands, Italy, Belgium, Slovakia, Spain
- America / Australia: USA, Canada, Australia, Argentina
- East Asia: Taiwan, South Korea, Singapore, Japan

Additional countries and regions (if necessary) may be added. If geographical region categories are used for stratified randomization, only the randomization strata will be used as covariate in respective analyses, in addition to the BMI category.

2.1.1 General definitions

Each patient in this study receives only one type of study drug, each of which is to be taken orally once daily. The term "study drug" therefore refers to LJN452 (10 µg, 30 µg, 60 µg, 90 µg, 140 µg or 200 µg), or matching placebo.

Date of first administration of study drug: Date of first administration as recorded in the CRF (Drug Administration Record page)

Date of last administration of study drug: Date of last administration as recorded in the CRF (Drug Administration Record page)

Study day: Study day is calculated from the date of first administration of study drug, which is defined as Day 1.

Baseline: Generally, baseline is defined as the last assessment before date and time of first administration of study drug; if only the date is available, the last assessment before or at the date of first administration of study drug will be used. For transaminases (ALT, AST, GGT) and bilirubin, the baseline value will be calculated as the mean of the last two assessments before first administration of the study drug, which are usually those taken at the Screening 1 and Baseline visits (Screening 2 and Baseline if a test was performed during Screening 2 visit).

Last contact: Date of last data point entered in the database.

Treatment period: Period from Baseline visit to End-of-treatment visit (included).

Follow-up period: Period from End-of-treatment visit (not included) to End-of-Study visit (included). Note that the follow-up period may consist of two follow-up epochs in some subjects.

2.2 Analysis sets

The following analysis sets will be defined for the statistical analysis:

- Screened set (SCR) – All subjects who signed the informed consent. Data from screen failures, as available, should be included in the analysis datasets even if not used for any CSR relevant analysis.
- Randomized set (RAN) – All subjects who received a randomization number, regardless of receiving trial medication.
- Full analysis set (FAS) – All subjects to whom study treatment has been assigned*. Following the intent-to-treat (ITT) principle, subjects are analyzed according to the treatment they have been assigned to at randomization.
- Safety set (SAF) - All subjects who received at least one dose of study drug and have at least one post-baseline safety assessment. Of note, the statement that a subject had no adverse events also constitutes a safety assessment. Subjects will be analyzed according to the treatment received. Of note, subjects with dose reduction due to AE will be analyzed according to the treatment they received up to the dose reduction. In case of dose administration errors, e.g. if a subject did not adhere to the intended medication schedule, subjects will be analyzed according to the following rules in the SAF, as far as they can be programmed based on recorded data:

1. If LJN452 was administered to a subject in the placebo group, the subject will be placed in the respective LJN452 dose group, independent of how long LJN452 was administered. Only the planned dose groups will be used (10, 30, 60, 90, 140 and 200 µg), so that a subject will be placed in the lowest dose group with daily dose equal or greater to the maximum daily dose that the subject received for at least one day.
2. If only placebo was administered to a subject in a LJN452 group, the subject will be placed in the placebo group.
3. If placebo was administered for a certain period to a subject in a LJN452 group, the subject will be placed in the respective LJN452 dose group. This refers in particular to subjects who took placebo erroneously for several days (e.g., 10 days) because they used capsules of a single bottle only instead of all three bottles.
4. If a higher than assigned dose of LJN452 was administered to a subject for at least 7 days (or an unknown period), the subject will be placed in the higher dose group. Same applies, accordingly, to cases where various doses of LJN452 were erroneously administered to a subject.

- If deviations from the assigned dose are unknown or cannot be derived from recorded data, the subject will be grouped into the assigned treatment arm.

* excluding subjects who were mis-randomized and did not take investigational drug. Mis-randomized subjects are those who were not qualified for randomization, but were inadvertently randomized into the study.

2.2.1 Subgroup of interest

The following subgroups of interest will be used for primary efficacy and safety analyses at the final analysis (subgroup analyses will be performed for the defined subgroup and its complement):

- NASH diagnosis confirmed by historic biopsy results (definition in 5.3.7) (only for Part A and Part B analyses at respective reporting events)
- BMI strata:
  1. non-obese (<30 kg/m^2 if Asian, < 35 kg/m^2 if Non-Asian)
  2. obese (≥ 30 kg/m^2 if Asian, ≥ 35 kg/m^2 if Non-Asian).

Additional subgroups analyses may be performed to assess the specificity of the selection criteria for the studied NASH population (for example, in patients with a history of diabetes of at least 5 years, or in those with NASH based on an algorithm [Bazick et al. (2015)]). However, given the sample size of the study, such additional analyses will only be considered if the subgroups and their complements are of a relevant size (e.g., at least 30% of the total sample size). Furthermore, these are not required for Part C of the study.

Additional subgroup tables of key outcomes will be produced by race and/or geographical region/country (e.g., Japan / Non-Japan in Parts B and C), as required.

Stratified and subgroup analyses will be based on actual (derived) BMI from eCRF data, not BMI recorded in the IRT data, unless the required eCRF data are missing.

2.3 Subject disposition, demographics and other baseline characteristics

Demographic variables and other baseline characteristics will be summarized for the FAS. In addition, all relevant medical history, and protocol solicited medical history will be summarized for the FAS. As protocol solicited medical history can be entered at both screening visits, only one value per subject resulting in Yes, No or missing will be derived (Yes > No > missing).

In addition to the baseline characteristics directly recorded in the eCRF or derived in the Oracle Clinical database, or loaded from third party databases, the following will be derived in the statistical database (see section 5.3 for derivation rules) and included in the summary of baseline characteristics:

- Diabetes status
- NAFLD fibrosis score
- Fibrosis stage (biopsy based)
- Fibrosis biomarker test
- [redacted]
- FIB-4
- [redacted]
- Diagnosis of NASH based on algorithm [Bazick et al. (2015)]
- Age category (<65 years, ≥65years)

The summary of demographic variables and other baseline characteristics will be produced for Part A alone, Part B alone, Part C alone and all parts combined.

2.3.1 Subject disposition

The number of subjects in each analysis set will be presented overall for the screened set and by treatment group for the randomized set.

The number and percentage of subjects in the randomized set who completed or discontinued each study period (screening, treatment, follow-up), and the reason for discontinuation will be presented for each treatment group and all subjects.

The frequency (%) of subjects with eCSR reportable protocol deviations as well as the criteria leading to exclusion from analysis sets will be presented in separate tables for the randomized set.

These summaries will be produced for Part A alone, Part B alone, and both parts combined.

2.4 Treatments (study treatment, rescue medication, concomitant therapies, compliance)

2.4.1 Study treatment / compliance

Overall duration (in weeks) for the double-blind investigational treatment will be computed for each randomized patient as follows:

(date of last administration of study drug – date of first administration of study drug + 1) / 7

If a patient was randomized, but did not receive any dose of randomized double blind study medication (protocol deviation) then the exposure is set to 0.

The duration of exposure to study drug (in weeks) will be summarized for the SAF, using descriptive statistics, and additionally by duration category in steps of "week":

- ≥0 week- < 1 weeks
- ≥1 week- < 2 weeks
- (…)
- ≥ 11 weeks - < 12 weeks
- ≥ 12 weeks - < 20 weeks
- ≥ 20 weeks - < 28 weeks
- ≥ 28 weeks - < 36 weeks
- ≥ 36 weeks - < 48 weeks
- ≥ 48 weeks

Note that 1-week intervals are only used up to 12 weeks. For Part C, this allows a summary in comparison to Parts A and B.

Furthermore, the proportion of subjects with dose reduction will be summarized for the SAF.

2.4.2 Prior, concomitant and post therapies

Medications will be identified using the WHO dictionary including ATC code and presented for the SAF. Prior medications are defined as any medications taken prior to the first administration of study drug (regardless of whether they are stopped or continued after randomization). Concomitant medications and significant non-drug therapies are defined as those used during the treatment and follow-up period (i.e., until end of study visit). Prior and concomitant medications will be summarized by treatment group in separate tables. Medications will be presented in alphabetical order, by ATC codes and grouped by anatomical main group (the 1st level of the ATC code). Tables will also show the overall number and percentage of subjects receiving at least one drug of a particular ATC code and at least one drug in a particular anatomical main group.

Concomitant medications that were prohibited as per protocol and given during the study participation as well as significant non-drug therapies will be provided in separate tables.

2.5 Analysis of the primary objective

2.5.1 Primary endpoint

The following are the primary variables and endpoints:

Safety (to be assessed in the SAF):

- Occurrence of SAE
- Occurrence of AE resulting in permanent discontinuation or dose reduction of study treatment

- Occurrence of AE of special interest

Efficacy (to be assessed in the FAS):

- Change from baseline to Week 12 in ALT
- Change from baseline to Week 12 in AST
- Relative change from baseline to Week 12 of percentage of liver fat

2.5.2 Statistical hypothesis, model, and method of analysis

No confirmatory statistical hypothesis testing will be performed in this study. The methods to analyze the primary safety and efficacy variables are outlined in Table 2-1.

Table 2-1 Primary variables and methods of analysis

Variable	Method of analysis
Occurrence of SAE	Summary table of absolute and relative frequency, overall and by preferred term
Occurrence of AE resulting in discontinuation or dose reduction of study treatment	Summary table of absolute and relative frequency, overall and by preferred term
Occurrence of AE of special interest	Summary table of absolute and relative frequency, overall and by type of AE (risk definition as per SPP)
Change from baseline to Week 12 of ALT	Baseline adjusted mean estimates and pairwise differences from repeated measures ANCOVA; descriptive statistics; dose-response modelling
Change from baseline to Week 12 of AST	Baseline adjusted mean estimates and pairwise differences from repeated measures ANCOVA; descriptive statistics; dose-response modelling
Relative change from baseline to Week 12 in percentage of fat in the liver assessed using MRI	Baseline adjusted mean estimates and pairwise differences from ANCOVA, descriptive statistics; dose-response modelling

Repeated measures ANCOVA models will include time (visit) and treatment group as categorical explanatory variables. The stratification factor (BMI group), geographical region and the baseline assessment will be included as covariates. At the Part A interim analysis, however, geographical region will not be included, due to the small sample size. BMI group may be dropped as well if necessary to obtain a model fit. The interaction terms of time with baseline assessment and treatment will be included as well. An unstructured covariance matrix will be assumed for the within subject repeated measurements, and Kenward-Rogers type degrees of freedom will be used. 95% confidence intervals will be calculated for treatment differences (without adjustment for multiple comparisons). Estimates will be derived for all time points up to Week 12. In Part A + B and A + B + C pooled analyses, as applicable, Week 1 data (only available in Part A) will not be used in the model.

For ALT and AST, a multiple contrast test to confirm a general trend over placebo will be performed in addition. A one-sided p-value ≤ 0.05 will be considered as a confirmation of a

dose-response relationship, without adjustment for multiple comparisons (due to testing more than one parameter). Optimal contrast vectors will be derived from pre-specified alternative dose-response shapes, and estimates and covariance matrix at Week 12 will be obtained from the reperated measures model. For this purpose, a random effects model with a random intercept (subject as a random effect) and same fixed effects as mentioned above will be used, but no specification of covariance structure and degrees of freedom are required (to enable running it in R/lme4, alternatively nlme/gls equivalent to specification above, without using Kenward-Roger adjustment if not available). In a second step, an averaged model will be fit to the data using a bootstrap procedure selecting the best model in each bootstrap sample based on Akaike Information Criterion (AIC). Estimates will be shown with 90% confidence intervals. A similar dose-response analysis will be performed for relative change from baseline to Week 12 of the percentage of liver fat. However, as there is only one post-baseline assessment, direct ANCOVA estimates will be used instead of the repeated measures estimates, which can be done in functions of the Dosefinding R package directly.

Graphical displays will be used as appropriate and will be defined in the specification of Tables, Figures and Listings (TFL shells). These will include, but not necessarily be restricted to:

- Line plots of ALT and AST mean change from baseline over time (estimates from repeated measures analysis)
- Spaghetti plots of ALT and AST change from baseline over time
- Individual subject profiles of liver enzymes over time (at least for Part A interim analysis)
- Plot of posterior probability of superiority over placebo for a range of differences, with respect to ALT and AST change from baseline at Week 8 (for Part A interim analysis) and Week 12, assuming a flat prior (therefore based on the likelihood estimation from observed data only)

2.5.3 Handling of missing values/censoring/discontinuations

Missing data for the liver function tests in Table 2-1 will be accounted for by the use of repeated measures ANCOVA (MMRM), assuming data are missing at random (MAR). In case of dose reduction or treatment discontinuation, any ALT and AST assessments after reduction / discontinuation will be set to “missing” for all efficacy analyses.

The repeated measures ANCOVA for ALT and AST will additionally be run on all obtained assessments, including those taken after any dose reduction or discontinuation.

The missing data pattern will be explored graphically at the final analysis. Missing data, in this context, also refers to assessments that were set to missing as described above. If the exploration raises concerns about deviation from the MAR assumption, and/or the proportion of missing data is large (e.g., >10%), the possible impact will be discussed.

Missing data for liver fat will be imputed by the baseline value for the Week 12 analysis. No imputation will be applied for the final analysis in Part C, where a MMRM model is used. In case of treatment discontinuation, assessments obtained more than 4 weeks after last treatment will be set to “missing”.

In general, in summaries by nominal visit follow-up visits will be included as recorded, regardless to the time interval between last study treatment and follow-up visit.

As there are no formal hypothesis tests and no confirmatory claims based on the results, no other alternative analyses are pre-planned.

2.5.4 Supportive analyses

Primary variables will also be summarized, using descriptive statistics, in the subgroups defined in 2.2.1. Primary efficacy variables will, in addition, be summarized for subjects who had a significant change of lifestyle during the study (diet, exercise), indicated by a weight loss of more than 10% from baseline to any post-baseline visit (versus subjects without such weight loss). These supportive analyses will only be included in the CSR as required, but will be part of graphical data exploration using, e.g., a ShinyApp.

Furthermore, primary efficacy variables will be analyzed excluding the subset of patients who have shown early spikes of ALT, defined as an increase from baseline by more than 100 U/L during the first two weeks after first administration of study drug (only applicable for Part A/B).

2.6 Analysis of the key secondary objective

There are no "key secondary" objectives defined in the protocol of this study.

2.7 Analysis of secondary efficacy objective(s)

2.7.1 Secondary endpoints

Table 2-2 Secondary efficacy variables and methods of analysis

Variable	Analysis
Absolute and relative change from baseline by visit in percentage of fat in the liver assessed using MRI	Baseline adjusted mean estimates and pairwise differences from ANCOVA (repeated measures for Part C), descriptive statistics
Weight BMI Waist-to-hip (WTH) ratio	Descriptive statistics by visit, including change from baseline, pairwise differences versus placebo with 95% CI from repeated measures ANCOVA (weight and BMI) or simple ANCOVA (WTH ratio)
FGF19 C4	Descriptive statistics (statset2) by visit, including change from baseline, pairwise ratio versus placebo with 95% CI from ANCOVA (ratio post-dose versus pre-dose (and versus baseline for C4) at week 6 back-transformed from log scale)
Liver stiffness (in kPa) by Fibroscan® Enhanced liver fibrosis panel (ELF) score Score of fibrosis biomarker test (originally known as Fibrotest®/ FibroSure®)	Descriptive statistics by visit (including change from baseline), pairwise differences versus placebo with 95% CI from ANCOVA (repeated measures for Part C)
GGT	Descriptive statistics by visit (including change from baseline), pairwise differences versus placebo with 95% CI from repeated measures ANCOVA

Variable	Analysis
Fasting lipids: Total cholesterol Trigylcerides LDL and HDL cholesterol LDL / HDL ratio Free glycerol Free fatty acids	Descriptive statistics (statset2) by visit (including %change and log transformed ratio to baseline), and pairwise ratio versus placebo with 95% CI from repeated measures ANCOVA (ratio to baseline back transformed from log scale) Additional covariate: use of lipid reducing concomitant medication (see 5.3).
At least a one point improvement in fibrosis (NASH CRN staging) without worsening of steatohepatitis at Week 48 compared to baseline	Descriptive statistics (absolute and relative frequency), differences, odds ratio and relative risk reduction versus placebo with 95% CI
At least a two point improvement in fibrosis (NASH CRN staging) without worsening of steatohepatitis at Week 48 compared to baseline	Descriptive statistics (absolute and relative frequency), differences, odds ratio and relative risk reduction versus placebo with 95% CI
Resolution of steatohepatitis without worsening of fibrosis (NASH CRN staging) at Week 48 compared to baseline	Descriptive statistics (absolute and relative frequency), differences, odds ratio and relative risk reduction versus placebo with 95% CI
Change of NAS from baseline to Week 48	Descriptive statistics

2.7.2 Statistical hypothesis, model, and method of analysis

Methods of analysis are listed in Table 2-2. Summary tables will be presented by treatment group and visit (as applicable) using descriptive statistics. All efficacy variables will be analyzed in the FAS and assessments obtained after dose reduction will be included.

Repeated measures ANCOVA will be performed as described in 2.5.2. Standard ANCOVA for change of liver fat will include baseline assessment and BMI stratification group as covariates, and treatment group as explanatory variable, with no interaction terms. Standard ANCOVA for ratio post-dose versus pre-dose at week 6 for FGF19 and C4 will include pre-dose assessment and stratification group as covariate and treatment group as explanatory variable, with no interaction terms. This analysis will be performed on log transformed data (ratio as well as pre-dose) and resulting estimates will be transformed back to provide adjusted geometric mean and 95% confidence intervals for the within-group and between-group estimates.

Binary biopsy based endpoints will be analyzed using logistic regression, including baseline fibrosis stage and BMI stratification group as covariates. See 5.3.11 for definitions.

Graphical displays will be used as appropriate and will be defined in the specification of Tables, Figures and Listings (TFL shells).

2.7.3 Handling of missing values/censoring/discontinuations

Missing data for the efficacy variables in Table 2-2 will be accounted for by the use of repeated measures ANCOVA (MMRM), as applicable, assuming data are missing at random (MAR). For variables with only one post-baseline assessment, a missing post-baseline value will be imputed by the baseline value. The same applies, correspondingly, to the pre-dose versus post-dose analysis of FGF19 and C4.

Missing follow-up biopsy results will not be imputed, so that the primary analysis of these parameters will only be based on subjects with valid paired biopsies who have received ≥ 24 weeks of the assigned treatment.

In addition, the following supportive analyses addressing different estimands will be performed for binary biopsy endpoints (response criteria):

- Non-responder imputation: all binary variables based on biopsy results will be considered as "non-response" if the value is missing and NAS score will be imputed by the baseline score in case of missing follow-up biopsy.
- Analysis corresponding to a hypothetical estimand: follow-up biopsy results for subjects who discontinue study treatment prematurely are discarded (set to missing), and missing results are imputed by multiple imputation (see 5.1.3.3).

2.8 Safety analyses

All safety analyses described in this section will be carried out using the Safety set (SAF). Patients will be analyzed according to the treatment they received.

All safety listings will indicate the actual treatment the patient was on during the treatment period.

2.8.1 Adverse events (AEs)

Treatment emergent adverse events (TEAE), defined as events that started after the first dose of study drug or were present prior to the first dose of study drug but increased in severity based on preferred term, will be summarized. All AE with onset until 28 days after last dose of study drug are considered treatment emergent.

The summary tables for these TEAE will present, for each treatment group, the number and percentage of subjects having experienced:

- any AE by primary system organ class (SOC) and preferred term (PT) (including "Any AE"),
- any SAE by primary system organ class (SOC) and preferred term (PT) (including "Any SAE"),
- any AE by preferred term (PT), which was reported in more than 5% of the subjects in any treatment group of Part B, or more than one subject in any of the treatment groups in Part A (can be produced manually for the CSR by cutting down the long table version)
- any AE by Standardized MedDRA Query (SMQ) and preferred term
- any AE belonging to Drug Related Hepatic Disorders – Comprehensive Search (SMQ) by preferred term
- any AE by SOC, PT and maximum severity,
- any AE possibly related to study treatment (investigator assessment) by SOC and PT
- any AE resulting in dose reduction or discontinuation of study treatment by SOC and PT,
- Incidence of potential cases of safety risks as defined in the safety profiling plan (SPP), including risks not based on AE records.

If a subject reported more than one adverse event with the same preferred term, the adverse event with the greatest severity will be presented. If a subject reported more than one adverse event within the same primary system organ class, the subject will be counted only once with the greatest severity at the system organ class level, where applicable.

To account for the longer duration of exposure to study treatment in Part C (48 weeks) compared to Parts A and B (12 weeks), exposure adjusted incidences for the types of events listed above will be presented additionally. The placebo groups from Parts A, B and C will be pooled for this type of analysis.

Summary reports will be based on the newest MedDRA version available at the time of generating the reports, the version will be specified in a footnote.

Listings will also be provided for SAEs that occurred in screening failures, and (by treatment group) for:

- SAEs
- AEs causing study drug discontinuation
- AEs requiring dose adjustment or interruption
- AEs suspected to be drug-related
- AEs defined as identified or potential risks

For the legal requirements of ClinicalTrials.gov and EudraCT, two required tables on treatment emergent adverse events which are not serious adverse events with an incidence greater than or equal to 5% (in any treatment group, Part A, Part B and Part C combined) and on treatment emergent serious adverse events and SAE suspected to be related to study treatment will be provided by system organ class and preferred term on the safety set population.

If for a same patient, several consecutive AEs (irrespective of study treatment causality, seriousness and severity) occurred with the same SOC and PT:

- *a single occurrence will be counted if there is ≤ 1 day gap between the end date of the preceding AE and the start date of the consecutive AE*
- *more than one occurrence will be counted if there is > 1 day gap between the end date of the preceding AE and the start date of the consecutive AE*

For occurrence, the presence of at least one SAE / SAE suspected to be related to study treatment / non SAE has to be checked in a block e.g., among AE's in a ≤ 1 day gap block, if at least one SAE is occurring, then one occurrence is calculated for that SAE.

The number of deaths resulting from SAEs suspected to be related to study treatment and SAEs irrespective of study treatment relationship will be provided by SOC and PT.

2.8.1.1 Adverse events of special interest / grouping of AEs

The number and percentage of subjects having experienced any event of special interest for LJN452 treatment will be summarized by class of identified or potential risk (as defined in the Safety Profiling Plan) and respective subcategories for each treatment group..

Definitions can be found in 5.2. Only treatment emergent events, as defined in 2.8.1, will be considered.

For the evaluation of pruritus, the time to resolution of pruritus will be analyzed using Kaplan-Meier analysis in subjects with pruritus. If the event has not yet resolved at the time of completion or the data extraction (for DMC and interim analysis) the observation will be censored at the End-of-study date/day or data cutoff date/day, respectively. Time to onset of pruritus will be analyzed with summary statistics in subjects with pruritus event. A pruritus event is defined as any event as defined in the respective eCRS category.

2.8.2 Deaths

The number and percentage of subjects who died will be summarized by primary cause of death for each treatment group.

2.8.3 Laboratory data

The summary of safety laboratory evaluations will be presented for the groups of laboratory tests (e.g., hematology, clinical chemistry). Descriptive summary statistics for the change from baseline to each study visit will be presented. These descriptive summaries will be presented by test group, laboratory test and treatment group. Change from baseline will only be summarized for subjects with both baseline and post baseline values. Relative and absolute frequencies of subjects with liver events as defined in 5.3.1 will also be provided, as well as shift tables based on the normal laboratory ranges. For the shift tables, the normal laboratory ranges will be used to evaluate whether a particular laboratory test value was normal, low, or high for each visit value relative to whether or not the baseline value was normal, low, or high. These summaries will be presented by laboratory test category and treatment group.

For ALT and AST, the following summaries will be shown in addition, by treatment group:

- Number and percentage of subjects with spikes (increase from baseline by more than 100 U/L) during the first two weeks after first administration of study drug.
- Maximum post-baseline value of each subject, in terms of change from baseline and multiple of upper limit of normal range.
- In subjects with ALT spikes, number and percentage of subjects unresolved and resolved with or without intervention at end of study.
- In subjects with ALT spikes, summary statistics for time to resolution of spikes (resolution defined as $\leq$ baseline + 25 U/L).

The summaries for ALT and AST spikes will not be produced for Part C of the study, but the "spike" flag will still be derived and included in the analysis datasets.

Safety laboratory parameters which are also part of the efficacy analyses will be included in the safety tables, too (enzymes, lipids, glucose). All available data will be used for safety summaries. For visits belonging to the treatment period in Part C (i.e., up to Week 48), assessments after discontinuation of study treatment will not be included in the summaries. Follow-up visits, however, will be included as reported, regardless of time interval between last study treatment and follow-up.

The number and percentage of subjects with clinically notable laboratory results after baseline will be presented. Clinically notable laboratory results, for those parameters where ranges are available, are presented in 5.3.1. The most extreme post-dose value will be considered. Only subjects with laboratory results at baseline and post-baseline from the central laboratory will be included in the tabulations.

In addition to summary tables, a listing of subjects with any treatment emergent notable lab results, defined as results obtained in the treatment period, will be provided by treatment group and subject number, along with their values at all visits, including repeated, unscheduled and follow-up lab values. In the listing, all abnormal values outside of normal range and notable values will be flagged, e.g., L/H for a value being below/above normal range, LN/HN for a value being in notable low/high range. Laboratory "Visit dates" from the visit panel on eCRF will not be listed, instead laboratory listings will present the date and day of the sample collection (obtained directly from lab panel). Comprehensive subject listings of all laboratory assessments (i.e., a data dump) will not be produced for the CSR.

Results from urine dipstick and routine urinalysis with Reflex Micro will not be reported in CSR outputs, but will be included in analysis datasets.

2.8.4 Other safety data

2.8.4.1 ECG and cardiac imaging data

ECG data will be summarized by treatment and visit. No cardiac imaging will be performed.

Notable QTc values and change from baseline will be summarized. A notable value is defined as a QTc interval of greater than 450 ms. The categories used for the change (increase) in QTc are - less than 30 ms, 30 to 60 ms and greater than 60 ms.

The Fridericia QT correction formula (QTcF) will be used for clinical decisions.

2.8.4.2 Vital signs

Analysis of the vital sign measurements using summary statistics for the change from baseline for each post-baseline visit will be performed. These descriptive summaries will be presented by vital sign and treatment group. Change from baseline will only be summarized for subjects with both baseline and post-baseline values. Subjects with notable vital signs as defined in 5.3.9 will be listed.

2.9 Pharmacokinetic endpoints

All subjects with valid PK data will be included in the pharmacokinetic (PK) data analysis. Plasma concentrations will be expressed in ng/mL. LJN452 plasma concentration data will be listed by part, cohort, treatment group, subject, and visit/sampling time point. Descriptive summary statistics will be provided by part, cohort, treatment group and visit/sampling time point. Summary statistics will include mean (arithmetic and geometric), SD, CV (arithmetic and geometric), median, minimum and maximum. All concentrations below the limit of quantification or missing data will be labeled as such in the concentration data listings. Concentrations below the Limit of Quantification (LLOQ) will be treated as zero in summary

statistics for concentration data only, and summary tables will also show the number and percentage of such values.

Any other PK analyses will be described in a separate report (e.g. a CSR addendum). For instance, a population PK modeling approach will be used to predict exposure. PK data from study CLJN452X2101 in healthy subjects will be used to build the PK model. Subsequently, a population PK model for PK data from this study (CLJN452A2202) will be established. If necessary, the combined data sets from the two studies will be used. Once an adequate model is established, to be verified by applying various model diagnostics tools, the exposure within a dosing interval will be calculated for each subject.

2.10 PD and PK/PD analyses

All subjects with valid PK and PD data will be included in the data analysis. [redacted]

Correlation between PD markers (FGF19, C4) and efficacy markers (ALT, GGT, fibrosis markers) will also be explored graphically (not necessarily to be included in the CSR Appendix).

[redacted]

2.11 Patient-reported outcomes

The impact of LJN452 on various aspects of patient's health status will be assessed by the following patient reported outcome (PRO) tools:

- VAS for Itch
- [redacted]

Table 2-3 Patient reported outcome tools and methods of analysis

Variable	Method of analysis
VAS for Itch	Descriptive statistics by visit (including change from baseline), pairwise differences versus placebo with 95% CI from repeated measures ANCOVA
[redacted]	[redacted]

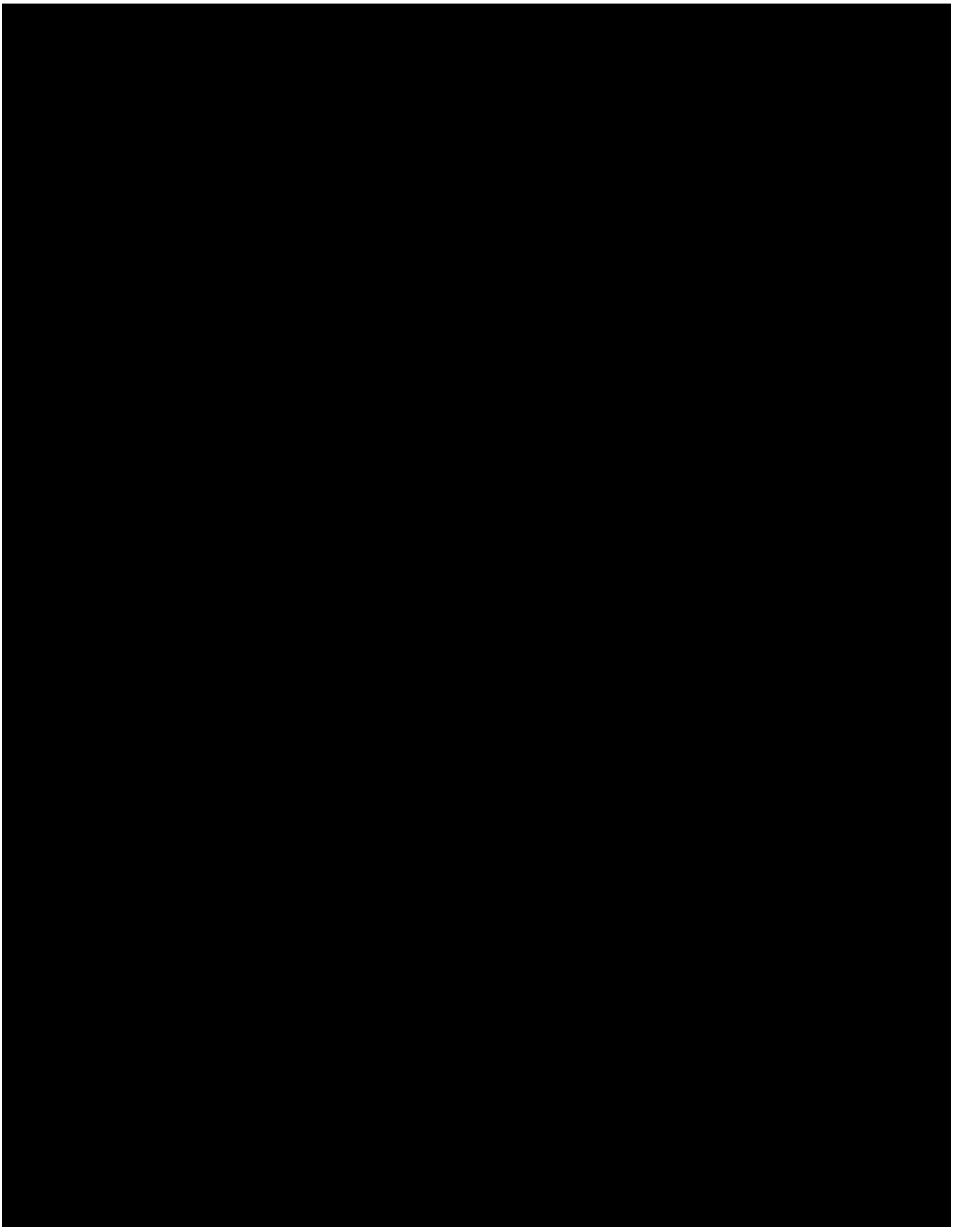
Variable	Method of analysis
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

Summary tables will be presented by treatment group and visit (as applicable) using descriptive statistics. All PRO variables will be analyzed in the FAS and assessments obtained after dose reduction will be included. With the exception of the VAS for Itch [redacted], PRO tools are only used in Part B and C of the study and therefore only Part B and C treatment groups will be displayed in the corresponding analyses.

[redacted]

Repeated measures ANCOVA will be performed as described in 2.5.2.

[redacted]



[redacted]

2.14 Interim analysis

An interim analysis will be conducted when ≥90% of the subjects in Part A have completed the Week 8 assessments. The analysis will, however, include all data available at the resulting data cutoff point. An independent DMC will review the safety profile and primary efficacy variables. Estimates for the efficacy variables (ALT, AST and GGT) will also include time points beyond Week 8 (using a repeated measures ANCOVA).

The following analyses will be performed:

- Demographic & background characteristics: as in 2.3
- Subject disposition: as in 2.3.1
- Duration of exposure to study drug, by treatment: as in 2.4.1
- ALT, AST and GGT over time: summary statistics by visit and treatment group, estimates of difference versus placebo in each group using a repeated measures model (as described in 2.5.2), dose-response testing and modelling for ALT change from baseline at Week 8 and Week 12
- Change from baseline in % hepatic fat at Week 12: summary statistics by treatment group (only for subjects with available data), and individual subject profiles
- Adverse events: as in 2.8.1 (including 2.8.1.1 and 2.8.2)
- Newly occurring or worsening notably abnormal laboratory values: as described in 2.8.3
- Urinalysis: as described in 2.8.3
- Vital signs: as in 2.8.4.2
- Pharmacokinetic data (listings of individual concentrations and predicted exposure as available)
- Individual lipid profiles over time
- Summary statistics for FGF19, C4, [redacted] if data are available

Interim data analysis will be performed by an independent statistician and programmer, using unblinded treatment group information, and provided to the DMC.

Based on the results of the IA, the DMC will make recommendations on the treatment groups to be studied in Part B of the study. There are no strict quantitative rules for the DMC decision and no hypothesis tests. In brief, the DMC will determine which doses in Part A are safe, and among the safe doses, which are efficacious based on biomarker results (primarily ALT and AST).

A benchmark for biomarker response is a decrease from baseline of at least 27 U/L for ALT (mean displayed in [Neuschwander-Tetri et al (2015)] for obeticholic acid at week 12). Target thresholds for desired biomarker response and further guidance may be provided in the DMC charter.

Up to two of the doses are planned to be selected for Part B. In the event that the DMC selects only one active dose (safe and efficacious) to be tested in 75 subjects in Part B, one of the other

originally planned active treatment arms (highest safe but inefficacious dose) will continue with a smaller sample size (20 subjects) to confirm the earlier findings of this treatment arm observed in Part A. If only one dose is deemed safe then only one experimental group will be studied in Part B.

If none of the doses is considered safe, or no sufficient efficacy is observed with safe doses, Part B may not be initiated. As there are no formal hypothesis test, an adjustment of the type I error is not necessary to account for this interim analysis.

A second interim analysis (End-of-Part A analysis) will be performed when all subjects of Part A have completed the Week 16 visit, including only the subjects randomized to Part A. This analysis will not be performed by an independent team. However, as Part A and Part B cohorts use separate sections of the randomization list, only Part A treatment allocation information will be made available to the statisticians and programmers at this stage. The analyses performed will be a subset of the analyses planned for the final CSR and will be defined in a separate tracking sheet.

In addition, the DMC will review safety, including AEs and laboratory parameters as decribed above, on a regular basis after start of Part B. The independent statistician and programmer will prepare these safety reports using semi-blinded treatment group information.

The standard tables and listings produced for safety DMC reviews are a subset of the final CSR tables and listings and are listed in the DMC charter. They will also be identified in the TFL shell documentation.

A full analysis of Part A and B data will be performed when all subjects of Part B have completed the Week 16 visit, including all subjects randomized to Part B, but not any data from subjects randomized to Part C. Therefore, Part C will remain blinded at this stage (process as described above for the End-of-Part A analysis).

An additional interim analysis will be conducted when all patients randomized into Part C have completed the Week 12 visit. The scope of this analysis will be similar to the analysis at end of Part A. The analysis will be performed by the regular study team, which will be unblinded to Part C treatment allocation at this stage. Investigators, site personnel and study subjects will remain blinded until completion of the study.

3 Sample size calculation

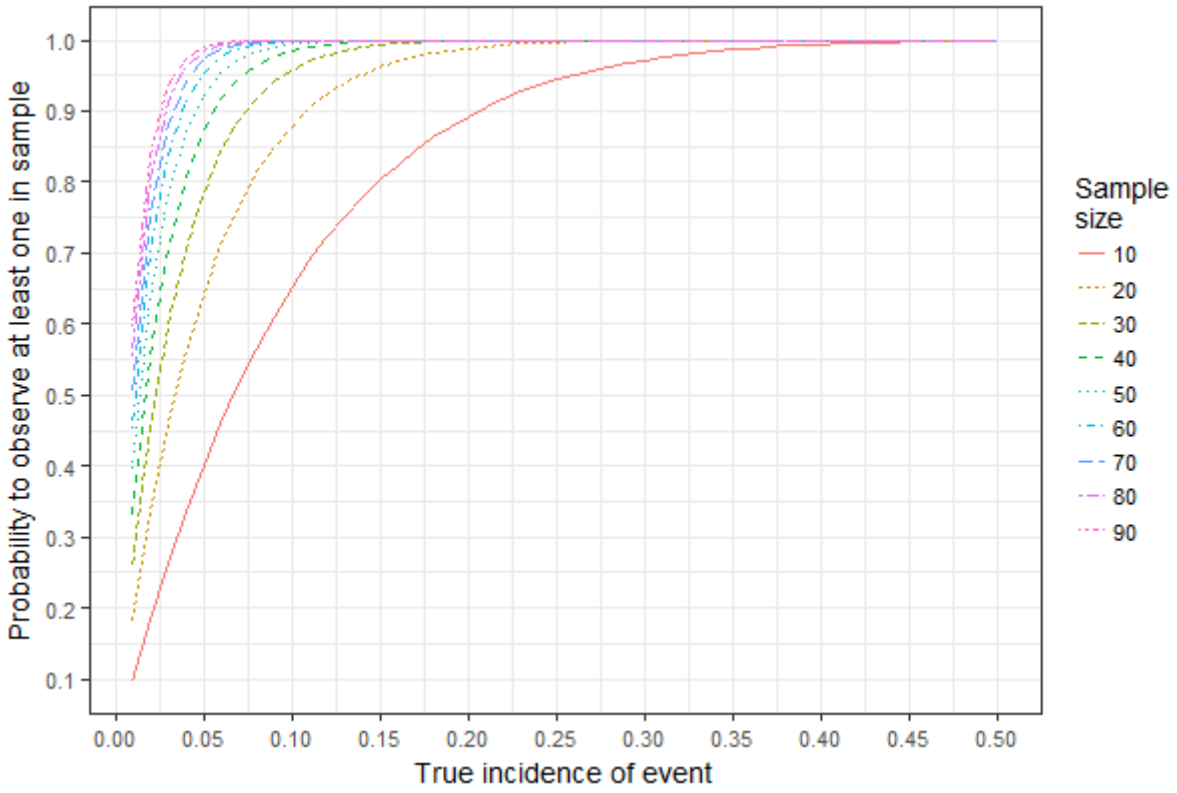
The primary objective of the study is to determine a safe dose or dose range. However, the assessment will be made based on the whole safety profile and not on quantitatively formulated hypotheses for distinct parameters. Therefore, the sample size is based on practicability with respect to expected speed of enrolment and duration of the study, not on formal statistical criteria.

3.1.1 Power considerations with given sample size for safety assessment

Events with a true incidence of 30% and above are observed with almost 100% probability in samples of 15 (Part A), 40 (combined Part A+B placebo) and 90 (combined Part A+B active dose) patients. This would include, for example, the isolated ALT elevations observed at high doses in healthy volunteers. Events with true incidences below 10% down to 3% are still very

likely to be observed in combined sample sizes from Parts A and B, and also in each dose in Part C , while events are observed with less than 50% probability in Part A if the true incidence is less than about 4% (Figure 3-1). It is noteworthy, however, that a single patient constitutes 6.7% in a sample of 15.

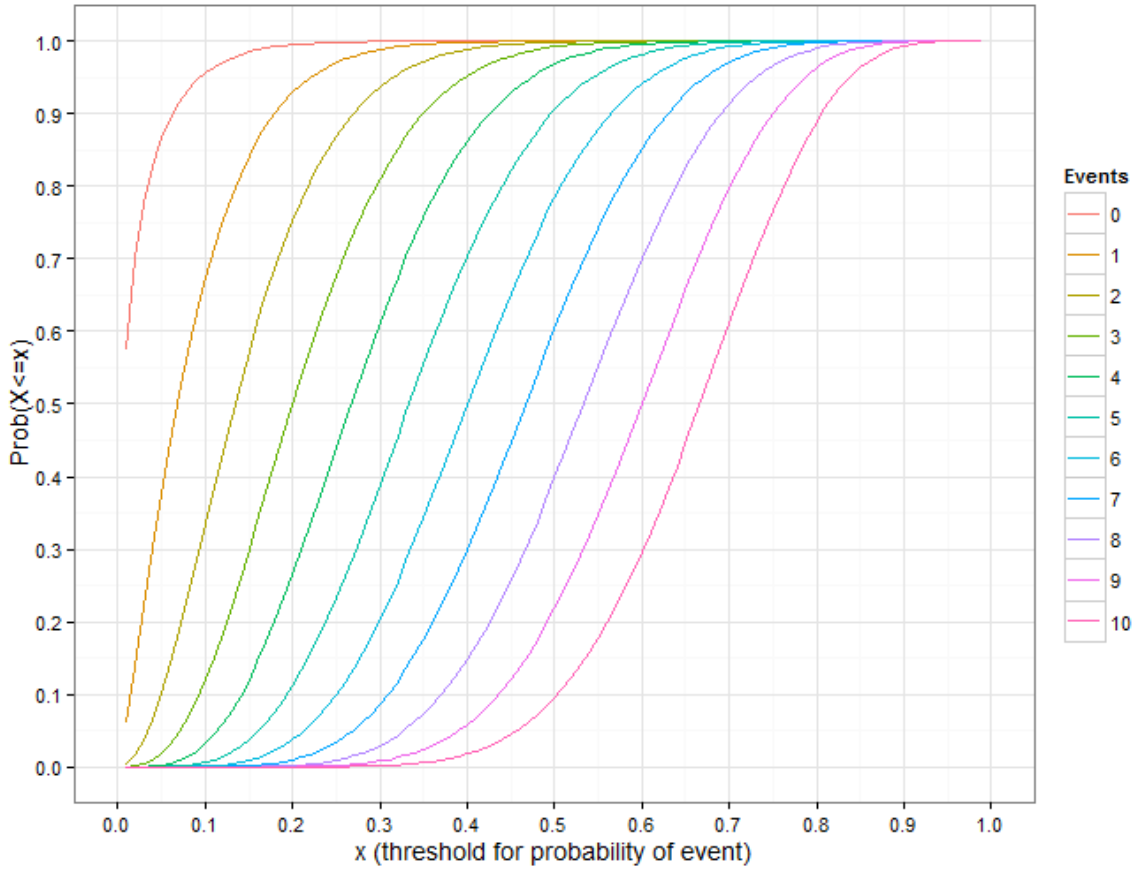
Figure 3-1 Binomial probability to observe an event with given sample size



IIn the interim analysis, predictions of the probability of an event (incidence) will be based on the observed number of events, to support the decision to continue or drop a dose for Part B.

Probabilities of the incidence being below a certain threshold are plotted in Figure 3-2 for a sample size of 15 patients when the event is observed in 0, 1, 2, … , 10 patients (assuming a beta distribution with prior shape parameters 0.33, 0.33).

Figure 3-2 Predictions for probability of event based on observed number



For example, if 0 events are observed, the probability that the incidence is ≤ 5% would be 87%. If an event is observed in one patient, the probability that the incidence is ≤ 5% would be 38%. Similarly, if an event is observed in 5 patients (one third), the probability that the incidence is ≤ 50% would be 90%, but the probability that the incidence is ≤ 30% would be 39% (calculated using R function pbeta).

3.1.2 Power considerations with given sample size for efficacy assessment

A consideration for primary efficacy analyses is given in the following.

[Neuschwander-Tetri et al (2015)] reported a mean change of ALT from baseline to week 12 of -28 for obeticholic acid (OCA) and -11 for placebo, with standard deviations of 48 and 33, respectively. Assuming for simplicity a common standard deviation of 45, this translates into an effect size (mean difference / standard deviation) of approximately 0.38, which can be considered as a benchmark. The power for a t-test to compare two groups (1-sided type I error 0.05) based on such an effect size would be 63% (59%) with a sample size of 90 (50) in the active and 40 (50) in the placebo group. If the active dose were slightly better than OCA (-33 versus -11), the power would be 81% (78%) Numbers in parentheses refer to comparisons in Part C with 50 subjects in the active arm(s) and 50 in the placebo arm where no pooling with subjects in the placebo arms of Parts A and B is considered (calculations with NQuery Advisor

7.0). Based on pre-clinical data for LJN452 and OCA, it is expected that the effect size achieved with an optimal dose of LJN452 maybe even larger, therefore resulting in a power above 80%.

For relative reduction of liver fat, no data are available for OCA. In the Novartis sponsored study CLCQ908A2216 in patients with NAFLD, with a baseline percentage of approximately 16%, the relative decrease after 12 weeks of treatment was about 2% in the placebo group and 21% in the highest active dose group, with a common standard deviation of approximately 30, resulting in an effect size of 0.63. If we assume a slightly smaller effect size of at least 0.5 for a LJN452A dose in the NASH study population, a power of ≥ 83% (79%) is achieved for pairwise comparisons with assumptions as described above.

Alternatively, we can consider the power of a multiple contrast test to demonstrate a trend over placebo across multiple dose arms, in this case 10 µg, 30 µg, 60 µg, 90 µg, 140 µg and 200 µg. We assume the dose-response curves in Figure 3-3. A beta shape is included as a possibility because the ALT elevations that may occur with high doses could, on average, interfere with the desired effect of ALT decrease.

Figure 3-3 Potential dose-response curves

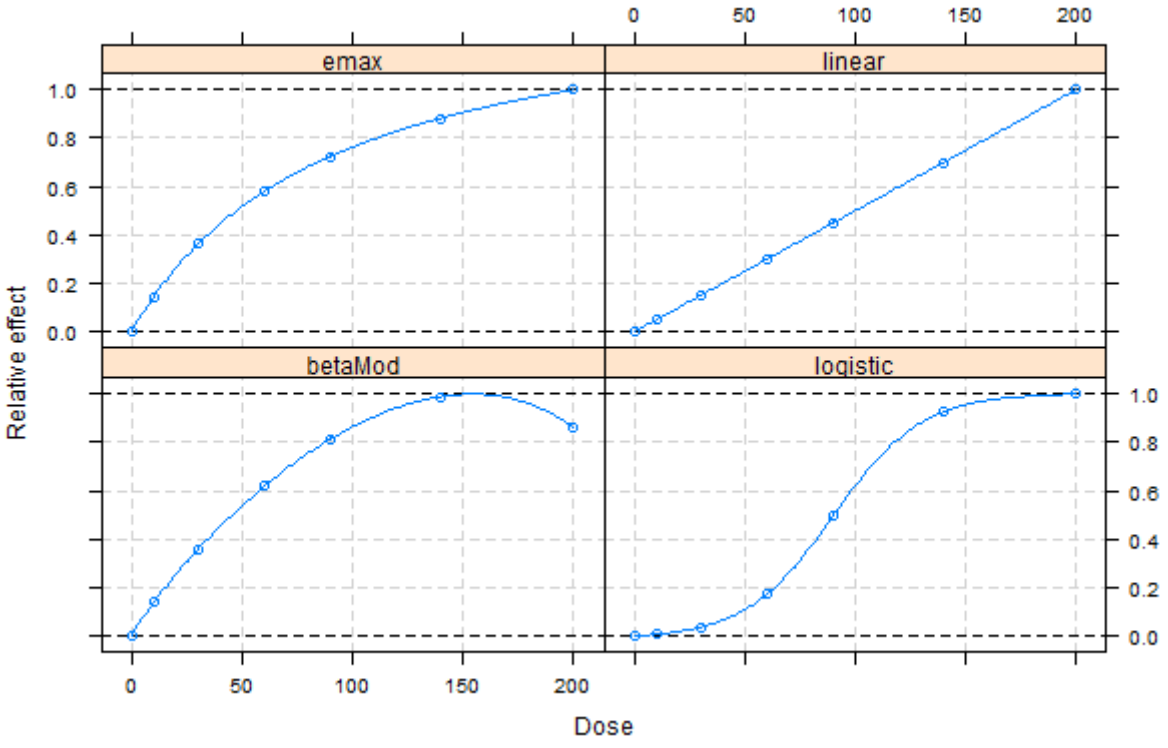


Table 3-1 Power for multiple contrast test for trend over placebo

Sample size in Doses (µg)							Power for model (%)				
0	10	30	60	90	140	200	Emax	Linear	Beta	Logistic	Average
90	15	15	35	90	50	50	82 (98)	75 (89)	83 (96)	85 (99)	81 (95)

Contrasts for multiple contrast test are optimal for each model type and sample size; assumptions: placebo response: -11, maximal response: -28 (in brackets: -33), common standard deviation: 45, type I error: 0.05 (one-sided);
Only Part B and C sample sizes after DMC recommendation were considered.

The average power for this type of test is at least 81% for a pooled Week 12 analysis of all study parts:

- 50 subjects assigned to two doses (140 and 200 µg) and 50 to placebo in Part C,

An average power of ≥ 95% is achieved if the effect size is moderately better than that of OCA (Table 3-4, calculations using powMCT function of DoseFinding package in R). Power calculations for a multiple contrast test on relative reduction of liver fat were not performed, but are expected to be in a similar range based on the effect size considerations above.

As the first interim analysis will be conducted with only 13-15 patients per arm who have completed the Week 8 assessments, the power to compare the effect on ALT or AST reduction between groups will be considerably lower. Furthermore, only a certain percentage of the patients will have data up to Week 12 available at that time point, depending on the speed of enrollment. The assessment of efficacious doses in the interim analysis will therefore be primarily based on the relative size of the point estimates (mean changes) and the shapes of the curves over time for the liver enzymes in each group.

3.1.3 Power consideration for biopsy endpoints in Part C

The longer treatment duration in Part C is due to the DMC recommendation to include paired biopsy assessments. Therefore, the sample size in Part C is partly based on the following assumptions for biopsy based outcomes:

- Response parameter: Achievement of at least one stage improvement of fibrosis with no worsening of steatohepatitis at Week 48 compared to baseline (binary).
- Placebo response rate: 13%, based on interpolated results from [Neuschwander-Tetri et al (2015)]. This assumes a linear improvement over time (a placebo rate of 19% was reported at week 72).
- Tropifexor response rate: 37% (best case assumption) with at least one of the doses.

For a 2-group continuity corrected χ^2 test of proportions with type error 0.05 (1-sided, no adjustment for multiple comparisons), a sample size of 50 per group results in a power of 82% (nQuery Advisor 7.0). The actual power might be smaller due to missing follow-up biopsies.

4 Change to protocol specified analyses

No relevant changes from protocol specified analyses were made.

5 Appendix

5.1 Imputation rules

5.1.1 Study drug

The date of the end-of-treatment visit (EOT, Week 12) will be used if the date of last adminstration of study drug is missing and the date of randomization will be used if the date of first administration is missing.

5.1.2 AE date imputation

Adverse Event Start Date Imputation (#IMPUTAEV):

This algorithm is expressed in the Variable Source Derivation column as **#IMPUTAEV(*event*)** where *event* is the partial start date of the adverse event.

The following table explains the notation used in the logic matrix. Please note that **missing start dates** will not be imputed.

	Day	Month	Year
Partial Adverse Event Start Date	Not used	MON	YYYY
Treatment Start Date	Not used	TRTM	TRTY

The following matrix explains the logic behind the imputation.

	MON MISSING	MON < TRTM	MON = TRTM	MON > TRTM
YYYY MISSING	NC Uncertain	NC Uncertain	NC Uncertain	NC Uncertain
YYYY < TRTY	(D) Before Treatment Start	(C) Before Treatment Start	(C) Before Treatment Start	(C) Before Treatment Start
YYYY = TRTY	(B) Uncertain	(C) Before Treatment Start	(B) Uncertain	(A) After Treatment Start
YYYY > TRTY	(E) After Treatment Start	(A) After Treatment Start	(A) After Treatment Start	(A) After Treatment Start

The following table is the legend to the logic matrix.

If AE end date is complete and AE end date < Treatment start date then AE start reference = min (informed consent date, earliest visit date from SV)

Else if AE end date is partial or AE is ongoing then AE start reference = treatment start date

Relationship	Date Imputation
Before AE Start reference	Partial date indicates AE start date **prior** to AE start reference
After AE Start reference	Partial date indicates AE start date **after** AE start reference
Uncertain	Partial date **insufficient to determine** relationship of AE start date to AE start reference
Imputation Calculation	
NC / Blank	No convention
(A)	01MONYYYY

(B)	Treatment start date+1
(C)	max(15MONYYYY, the start date of the screening period +1 day)
(D)	max(01JULYYYY, the start date of the screening period +1 day)
(E)	01JANYYYY
Complete date	No date imputation

Adverse Event End Date Imputation:

Imputed date = date part of original date, if complete date

If a patient is not randomized, then the AE end date is the date of completion of pre-randomization period.

If a patient is randomized:

- If AE end date, month, and year are missing or just month is missing, then the AE end date is set to the study completion/discontinuation visit date.

- If AE day is missing, then it is set to minimum of (treatment end day, last day of the month).

If imputed AE end date is less than the AE start date, use the AE start date as the imputed AE end date.

Impute Date Flag:

If year of the imputed date <> YYYY then date flag = Y

else if month of the imputed date <> MON then date flag = M

else if day of the imputed date <> day of original date then date_flag = D

else date flag = null

5.1.3 Concomitant medication date imputation

This algorithm is used when *event* is the partial start date of the concomitant medication or non-drug therapy/procedure.

The following table explains the notation used in the logic matrix. Please note that **completely missing start dates** will not be imputed. Also note that imputation of a start date must not result in a date later than the end date. In such case, start date will be max(01-MMM-YYYY, Treatment start date) if only the day is missing and max(01-JAN-YYYY, Treatment start date) if day and month are missing.

	Day	Month	Year
Partial CMD Start Date	Not used	MON	YYYY
Treatment Start Date	Not used	TRTM	TRTY

The following matrix explains the logic behind the imputation.

	MON MISSING	MON < TRTM	MON = TRTM	MON > TRTM

YYYY MISSING	**(C2)** Uncertain	**(C1)** Uncertain	**(C1)** Uncertain	**(C1)** Uncertain
YYYY < TRTY	**(D)** Before Treatment Start	**(A)** Before Treatment Start	**(A)** Before Treatment Start	**(A)** Before Treatment Start
YYYY = TRTY	**(C2)** Uncertain	**(A)** Before Treatment Start	**(C2)** Uncertain	**(B)** After Treatment Start
YYYY > TRTY	**(E)** After Treatment Start	**(B)** After Treatment Start	**(B)** After Treatment Start	**(B)** After Treatment Start

The following table is the legend to the logic matrix.

Relationship	**Date Imputation**
Before Treatment Start	Partial date indicates CMD start date **prior** to Treatment Start Date
After Treatment Start	Partial date indicates CMD start date **after** Treatment Start Date
Uncertain	Partial date **insufficient to determine** relationship of CMD start date relative to Treatment Start Date
Imputation Calculation	
(A)	15MONYYYY
(B)	01MONYYYY
(C1 or C2)	IF relative reference start = before THEN Treatment Start Date -1 day ELSE IF relative reference start = ' ' THEN Treatment Start Date +1 day
(D)	01JULYYYY
(E)	01JANYYYY

Concomitant Medication End Date Imputation:

If not ongoing then -

Imputed date = date part of CMENDTC, if complete date

Imputed date = min(reference end date, DEC 31) , if month is missing, (C2, D, E)

Imputed date = min(reference end date, last day of the Month) , if day is missing. (A, B, C1)

Concomitant Medication Date Flag:

If not a complete date then

Y - If year of the imputed date <> YYYY else

M – If month of the imputed date <> MON else

D

5.1.3.1 Prior therapies date imputation

Same as concomitant medication date imputation (as applicable).

5.1.3.2 Post therapies date imputation

Not applicable.

5.1.3.3 Multiple imputation of missing biopsy outcomes

A multiple imputation approach will be applied to missing follow-up biopsies in a supportive analysis. It is assumed that biopsy results are missing at random (MAR). Available results for subjects who discontinued study drug prior to 24 weeks of exposure (but still have a biopsy read-out) are also set to missing for this analysis. Imputation will be based on available results from subjects who did not discontinue study treatment, thus addressing the question what the results would have been if subjects had not discontinued. The imputation model will take into account the treatment group, the baseline fibrosis stage (categorical variables) and the baseline NAS score (continuous variable). As there is only one post-baseline biopsy assessment, an arbitrary missing pattern is assumed and a fully conditional specification (FCS) method with logistic regression is applied. The same logistic model is used to analyze the multiple imputed datasets, and pooled results are obtained based on Rubin's combination rules. The implementation can be done in SAS procedures MI (FCS LOGISTIC statement) and MIANALYZE. It is suggested to use 50 imputations.

5.1.3.4 Other imputations

Not applicable.

5.2 AEs coding/grading

AEs are coded using the MedDRA dictionary.

AE severity is classified using CTC grades if available. Grades represent 1=mild, 2=moderate, 3=severe, 4=life-threatening. If CTC grades are not available, these respective categories are collected in the eCRF.

AEs of special interest

AEs of special interest are those classified as identified or potential risks in the Safety Profiling Plan. In addition to risks based on AE terms, the following risks will also be tabulated at least for the final CSR (criteria to be defined in eCRS or separate document; list may be updated without amending this SAP):

- QTc prolongation
- Liver injury (lab data)
- Effects on lipid parameters (lab data)
- Increased ALT (lab test only)
- Renal effects (effects on renal function,) (lab data)

5.3 Derivations

5.3.1 Laboratory parameters

Liver events defined by laboratory parameter abnormalities (additional non-lab criteria as provided in the protocol are ignored):

- ALT or AST > 5 × ULN
- ALP > 2 × ULN

- TBL > 2 × ULN
- ALT or AST > 3 × ULN and INR > 1.5
- ALT or AST > 3 × ULN and TBL > 2 × ULN AND ALP to ≤ 2 × ULN (Potential Hy's Law cases)
- ALT or AST > 3 × ULN

Criteria for other laboratory parameters:

Table 5-1 Notable criteria for other laboratory parameters

Parameter	Threshold value	Unit
Albumin	<32	g/L
Hemoglobin	<70	g/L
Hemoglobin	>200	g/L
White blood cell count	<2.0	10^9/L
White blood cell count	>35.0	10^9/L
Platelets	<50	10^9/L
Platelets	>1000	10^9/L
Prothrombin Time INR	>4.0	
PT	>40.0	sec
APTT	>80.0	sec
Sodium	<120	mmol/L
Sodium	>160	mmol/L
Potassium	<3.0	mmol/L
Potassium	>6.0	mmol/L
Glucose	<2.2	mmol/L
Glucose	>27.8	mmol/L
Calcium	<1.50	mmol/L
Calcium	>3.00	mmol/L
Phosphate	<0.29	mmol/L
Creatinine	>177	µmol/L
Calculated eGFR	<60	mL/min

[redacted]

5.3.3 Fibrosis biomarker test, originally called Fibrotest®/ Fibrosure®

The score is calculated as:

$z = 4.467 \times \log_{10}(\alpha 2\text{-macroglobulin}) - 1.357 \times \log_{10}(\text{Haptoglobin}) + 1.017 \times \log_{10}(\text{GGT}) + 0.0281 \times \text{Age} + 1.737 \times \log_{10}(\text{Bilirubin}) - 1.184 \times \text{ApoA1} + 0.301 \times \text{Sex (0=female, 1=male)} - 5.540.$

Where:
α2-macroglobulin is given in g/L,
Haptoglobin is given in g/L,
GGT is given in U/L,
Age (at baseline) is given in years,
Bilirubin is given in µmol/L,
ApoA1 is given in g/L,
Sex is given as 0 for female and 1 for male.

Reference: [http://www.google.com/patents/US6631330]

5.3.6 Derived baseline characteristics

5.3.6.1 Diabetes

The diabetes status at baseline is determined as follows:

Diabetes = Yes if:

- Type 1 diabetes mellitus is ticked "Yes" in Protocol Solicited Medical History OR
- Type 2 diabetes mellitus is ticked "Yes" in Protocol Solicited Medical History OR
- Baseline fasting glucose > 5.6 mmol/L (100 mg/dL).

Otherwise Diabetes = No.

5.3.6.2 Use of lipid reducing drugs (e.g. statins)

Concomitant use of statins is determined as follows:

If drug belonging to ATC code C10 was used either starting prior to randomization and ongoing at randomization, or starting during the treatment epoch, concomitant use of statins is considered Yes, otherwise No.

5.3.7 NASH diagnosis by historical biopsy

Diagnosis of NASH by historical biopsy (as opposed to phenotypic NASH) is determined by all of the following:

- Liver biopsy CRF page with date of assessment 2 years or less before randomization date,
- Fibrosis stage not indicating cirrhosis:
  Batts-Ludwig score not 4,
  Ishak score not F6,
  Knodell score not 4,
  Metavir score not 4,
  Kleiner-Brunt score not 4
- Diagnosis "NASH" in the liver biopsy CRF page.

5.3.8 Disease scores and diagnostic algorithms

5.3.8.1 NAFLD fibrosis score

The NAFLD fibrosis score is calculated as [Angulo et al. (2007)]:

z = -1.675 + 0.037 * age + 0.094 * BMI + 1.13 * diabetes (0=no, 1=yes) + 0.99 * (AST/ALT) – 0.013 * platelets – 0.66 * albumin,

where age is given in years, BMI in kg/m^2, platelets in $10^9/L$, albumin in g/dL, and age and diabetes (see 5.3.6.1) are referring to the baseline condition.

5.3.8.2 Algorithm for diagnosis of NASH

A diagnostic algorithm was developed by [Bazick et al. (2015)]:

Logit(P) = 27.00 + 0.106 * BMI (kg/m2) – 0.035 * waist (cm) + 0.068 * AST (U/L) – 0.016 * ALT (U/L) + 0.71 * albumin (g/dL) + 0.24 * HbA_{1c} (%) + 0.0570 * [redacted] + 0.0014 * ferritin (ng/dL) + 0.57 * white (0=no, 1=yes).

A subject is classified as having NASH if the calculated probability P is ≥ 0.77.

[redacted]

5.3.10 Vital signs

Table 5-2 Notable abnormalities in vital signs

Vital signs		Notable abnormalities	
		Absolute	Relative to baseline
Pulse rate (beats/min)		> 130	≥ 120 and increase from baseline ≥ 15
		< 40	≤ 50 and decrease from baseline ≥ 15
Blood pressure (mmHg)	Systolic	> 200	≥ 180 and increase from baseline ≥ 20
		<75	≤ 90 and decrease from baseline ≥ 20
	Diastolic	> 115	≥ 105 and increase from baseline ≥ 15
		< 40	≤ 50 and decrease from baseline ≥ 15

5.3.11 Definitions for biopsy based endpoints

5.3.11.1 Improvement of fibrosis compared to baseline

The determination of fibrosis improvement will be based on NASH CRN staging. Only main stages (0, 1, 2, 3, 4) will be considered. For example, a change from 1c to 1b or 1a will not be counted as a one point change.

Table 5-3 Fibrosis stages and possible outcomes

Baseline	Week 48	1 point improvement	2 point improvement
2	0	Yes	Yes

Baseline	Week 48	1 point improvement	2 point improvement
2	1	Yes	No
2	≥2	No	No
3	≤1	Yes	Yes
3	2	Yes	No
3	≥3	No	No

5.3.11.2 Worsening of steatohepatitis at Week 48 compared to baseline:

Several outcome variables will be derived to allow comparison with published data:

- Total NAS Score at Week 48 greater than at Baseline
- Any component score (Steatosis, Lobular inflammation or Hepatocyte ballooning) at Week 48 greater than at Baseline (definition in FDA draft guidance) – used for primary estimand
- Any component score of Lobular inflammation or Hepatocyte ballooning at Week 48 greater than at Baseline OR component score of Steatosis at Week 48 more than 1 point greater than at Baseline (definition in EMA reflection paper)

5.3.11.3 Resolution of steatohepatitis at Week 48 compared to baseline

Several outcome variables will be derived to allow comparison with published data:

Subjective definition:

- Resolution of steatohepatitis will be determined as diagnostic category "not NAFLD" or "NAFLD, not NASH" as provided in the central biopsy report.

Based on NAS:

- Lobular inflammation ≤ 1 AND Hepatocyte ballooning = 0 AND any value for steatosis (definition in FDA draft guidance as well as EMA reflection paper) – used for primary estimand
- Lobular inflammation ≤ 1 AND Hepatocyte ballooning ≤1 AND any value for steatosis (Novartis proposal to account for noise in assessment of balloning)

5.4 Statistical models

5.4.1 Primary analysis

There is no formal hypothesis testing in this exploratory study.

Repeated measures models for continuous dependent variables will be analyzed using SAS PROC MIXED, assuming an unstructured covariance matrix and Kenward-Roger type degrees of freedom.

The general model for the response vector of patient $i = 1,\ldots,n$, $\mathbf{y}_i = (y_{i1}, \ldots, y_{im_i})'$, is:

$$\mathbf{y}_i \sim N_{m_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i)$$

where $\mathbf{y}_i$, are stochastically independent, $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the d-dimensional Normal distribution and

$y_{i1}, \ldots, y_{im_i}$ are repeated measures in time

$\mathbf{X}_i$ $m_i \times p$-matrix of covariates

$\boldsymbol{\beta}$ p-dimensional vector of parameters

$\boldsymbol{\Sigma}_i$ $m_i \times m_i$-covariance matrix

The covariates are treatment (categorical), visit (categorical), the baseline value of the response variable, and the interactions between treatment and visit and baseline value and visit. In addition, the stratification factor (BMI group) and geographical region will be included as categorical covariates.

Estimates of the treatment effect within groups and differences between groups will be obtained using the LSMEANS and ESTIMATE statements, respectively, for individual time points. 95% confidence intervals will be calculated.

Model assumptions will be checked based on diagnostic plots, but as there are no formal hypothesis tests, the possible impact of questionable models will only be described.

The multiple contrast test to confirm a general trend over placebo for ALT and AST changes from baseline to week 12 will be performed using the R package "DoseFinding" according to the following approach:

1. Fit a MMRM model for the response as decribed above (using SAS PROC MIXED), or alternatively a random intercept model with fixed effects as above and subject as random factor, using R/lme4 or nlme (specification of covariance structure of the R matrix and Kenward-Roger adjustment of degrees of freedom are not important for this purpose).
2. Extract estimates and variance-covariance matrix at Week 12 from fitted model.
3. Perform multiple contrast test with pre-defined model types and optimal contrasts, using weights according to the sample sizes of each group (MCTtest function).
4. Draw bootstrap samples (at least 1000) and fit each of the pre-defined models to the data in each sample. Select the model with the best fit based on AIC and predict the dose-response curve with this model in each sample.
5. Derive median and other quantiles for predicted response over the dose range from the bootstrap samples. This will result in an averaged model.

The pre-defined model types corresponding to the contrast vectors are:

- Emax model f(d, θ)=E0+Emax*d/(ED50 + d) with ED50 = 10
- Linear model f(d, θ)=E0+δ*d
- Beta model f(d, θ)=E0+Emax*B(δ1, δ2)*(d/scal)^ δ1*(1-d/scal)^ δ2 with δ1 = 0.9, δ2 = 0.9
- Logistic model f(d, θ)=E0+Emax/(1+exp((ED50-d)/δ)) with ED50 = 40 and δ = 10

The shapes of these models are displayed in Figure 3-3.

In brief, for each candidate model a contrast test statistic, based on a linear combination of the treatment estimates per dose will be derived. The contrast coefficients will be chosen to maximize the power to detect the pre-specified candidate models. The global test decision is based on the maximum of the contrast test statistics. A critical value q controlling the type I error rate can be derived from the fact that the four test statistics approximately follow a multivariate normal distribution and the distribution of the maximum of a multivariate normal distribution. If the maximum contrast test statistic exceeds the critical value q, the overall null hypothesis of a constant dose-response curve is rejected. In practice, a one-sided p-value ≤ 0.05 will be considered as a confirmation of a dose-response relationship. See [Pinheiro et al (2006)].

5.4.2 Key secondary analysis

There are no "key secondary" objectives defined in the protocol of this study. For repeated measures models used in other secondary analyses, see 5.4.1.

5.5 Rule of exclusion criteria of analysis sets

Table 5-4 Protocol deviations and other conditions that cause subjects to be excluded

Deviation ID	Description of Deviation	Exclusion from
OTH12	Patient was rescreened but did not sign a new ICF	All analysis sets
OTH14	ICH-GCP non-compliance of study site with impact on data quality	All analysis sets
TRT04	No drug taken after randomization	SAF
	ICF not signed	All analysis sets

5.6 Other statistical aspects

5.6.1 Crude incidence and related risk estimates

For n patients each at risk to experience a certain event with probability π, the crude incidence is estimated as p=x/n, where x is the number of patients with the event.

Odds ratio and 100*(1-α)% confidence interval

For an investigational drug group with n_1 patients at risk, independent from placebo with n_0 patients at risk, of whom x_1 and x_0 experience a certain event with probability π_1 and π_0 respectively, the odds ratio is estimated as

$\frac{p_1/(1-p_1)}{p_0/(1-p_0)}$ with $p_1 = x_1/n_1$ and $p_0 = x_0/n_0$. A conditional exact 100*(1-α)% confidence interval will be obtained by using the SAS procedure PROC FREQ with statement EXACT OR.

Risk difference and 100*(1-α)% confidence interval

For an investigational drug group with n_1 patients at risk, independent from placebo with n_0 patients at risk, of whom x_1 and x_0 experience a certain event, the risk difference is estimated

as p_1-p_0 with p_1= x_1/n_1 and p_0=x_0/n_0. Exact unconditional confidence limits for the risk difference will be obtained with SAS procedure PROC FREQ and option RISKDIFF in the TABLES statement, specifying the RISKDIFF option also in the EXACT statement.

Geometric mean and coefficient of variation

The geometric mean will be presented for the baseline values, absolute post-dose values and for the ratio to baseline (or pre-dose, respectively) values. The geometric mean of the ratio to baseline will be presented in terms of % change from baseline and will be calculated as follows: (exp (mean of the log-transformed ratio to baseline values) -1)*100.

The Coefficient of Variation (CV) will be calculated for the baseline values, the absolute post-dose values and the ratio to baseline (or pre-dose, respectively) values.

6 References

Angulo P, Hui JM, Marchesini G, et al. The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. Hepatology 2007; 45:846-54.

Bazick J, Donithan M, Neuschwander-Tetri BA, et al. Clinical Model for NASH and Advanced Fibrosis in Adult Patients With Diabetes and NAFLD: Guidelines for Referral in NAFLD. Diabetes Care 2015; 38:1347-55.

Neuschwander-Tetri BA, Loomba R, Sanyal AJ, et al. Farnesoid X nuclear receptor ligand obeticholic acid for non-cirrhotic, non-alcoholic steatohepatitis (FLINT): a multicentre, randomised, placebo-controlled trial. Lancet 2015; 385(9972):956-65.

Pinheiro J, Bornkamp B, Bretz F. Design and analysis of dose-finding studies combining multiple comparisons and modeling procedures. J Biopharm Stat 2006; 16:639-56.