TRIAL STATISTICAL ANALYSIS PLAN

c38524321-02

BI Trial No.:	1404-0043
Title:	Multicenter, double-blind, parallel-group, randomised, 48 weeks, dose-ranging, placebo-controlled phase II trial to evaluate efficacy, safety and tolerability of multiple subcutaneous (s.c.) doses of BI 456906 in patients with non-alcoholic steatohepatitis (NASH) and fibrosis
Investigational Product(s):	BI 456906
Responsible trial statistician(s):	Trial Statistician [redacted] Tel.: [redacted] Email: [redacted]
Date of statistical analysis plan:	19JAN2024 SIGNED
Version:	2.0

1. TABLE OF CONTENTS

LIST OF TABLES

2. LIST OF ABBREVIATIONS

Term	Definition / description
ADA	Anti-drug antibodies
AE	Adverse Event
AESI	Adverse Event of Special Interest
ALT	Alanine aminotransferase (SGPT)
ANCOVA	Analysis of covariance
AP	Alkaline phosphatase
APRI	AST to Platelet Ratio Index
AST	Aspartate aminotransferase (SGOT)
ATC	Anatomical Therapeutic Chemical
BI	Boehringer Ingelheim
BMI	Body Mass Index
BP	Blood pressure
CAP	Control Attenuation Parameter
CI	Confidence interval
CLDQ	Chronic Liver Disease Questionnaire
C-SSRS	Columbia – Suicide Severity Rating Scale
CTX-III	C-terminal crosslinked telopeptide of type III collagen
eCRF	electronic Case report form
CTC	Common Terminology Criteria
CTP	Clinical Trial Protocol
CTR	Clinical Trial Report
CV	Arithmetic coefficient of variation
DBL	Database Lock
DMC	Data Monitoring Committee
DTA	Data Transfer Agreement
ECG	Electrocardiogram
EDMS	Electronic Document Management System
eGFR	Estimated Glomerular Filtration Rate
ELF	Enhanced Liver Fibrosis
EOT	End of treatment

Term	Definition / description
EQ-5D 5L	European Quality of Life Questionnaire – 5 Dimensions 5 Levels
FAS	Full Analysis Set
FFA	Free Fatty Acids
FGF-21	Fibroblast Growth Factor 21
Fib-4	Fibrosis-4 index
GGT	Gamma-glutamyl transferase
HDL	High Density Lipoprotein
HMW	High Molecular Weight
HR	Heart rate
ICH	International Conference on Harmonisation
IPD	Important protocol deviation
IRT	Interactive Response Technology
iSAT	Independent safety analysis team
iSTAT	Independent statistician
LDL	Low Density Lipoprotein
LLOQ	Lower Limit of Quantification
MAR	Missing at random
MCP-Mod	Multiple Comparison Procedures and Modelling
MedDRA	Medical Dictionary for Regulatory Activities
mg	milligram
MMRM	Mixed model for repeated measures
MNAR	Missing not at random
MRI-PDFF	Magnetic Resonance Imaging - Proton Density Fat Fraction
NA	Not applicable
NAb	Neutralising antibodies
NAFLD	Non-Alcoholic Fatty Liver Disease
NAS	NAFLD Activity Score
NASH	Non-Alcoholic SteatoHepatitis
PD	Protocol deviation
PK	Pharmacokinetics
PPS	Per protocol set
PRO	Patient Reported Outcome

Term	Definition / description
Pro-C3	N-terminal propeptide of type III collagen
PT	Preferred term
Q1	Lower quartile
Q3	Upper quartile
REP	Residual effect period
SAP	Statistical analysis plan
SD	Standard deviation
SOC	System organ class
TMF	Trial Master File
TS	Treated Set
UACR	Urine albumin/creatinine ratio
ULN	Upper limit of normal
ULOQ	Upper Limit of Quantification
VLDL	Very Low Density Lipoprotein
WHO	World Health Organisation

3. INTRODUCTION

As per ICH E9, the purpose of this document is to provide a more technical and detailed elaboration of the principal features of the analysis described in the protocol, and to include detailed procedures for executing the statistical analysis of the primary and secondary variables and other data.

This TSAP assumes familiarity with the Clinical Trial Protocol (CTP), including Protocol Amendments. In particular, the TSAP is based on the planned analysis specification as written in CTP Section 7 “Statistical Methods and Determination of Sample Size”. Therefore, TSAP readers may consult the CTP for more background information on the study, e.g., on study objectives, study design and population, treatments, definition of measurements and variables, planning of sample size, randomisation.

This document describes the main analysis of study 1404-0043. One interim analysis was planned. This was described in a separate interim statistical analysis plan, and is not addressed further in this document. The interim analysis results were not intended to be used to stop the trial (which continued without modification after the interim analysis was performed), nor to make confirmatory claims. No statistical adjustments to Type 1 error are therefore considered necessary.

This TSAP is based on the 1404-0043 CTP (Version 6.0, 27JUL2023), together with the latest versions of the eCRF and relevant Data Transfer Agreements (DTAs) for external vendor data. The local protocol amendment for China (Version 1.0, 20APR2022) is also relevant to TSAP details regarding handling of biopsy data.

All statistical analyses defined in this TSAP will be performed by the trial statistician (TSTAT) with support from other trial team members following unblinding of the trial at final database lock (DBL). This TSAP will be finalized and signed prior to final DBL.

A DBL meeting will be held prior to taking the final database snapshot. It will be decided at this meeting whether any open cleaning issues will be accepted or not. Important protocol deviations (IPDs) will also be finalized prior to final DBL.

SAS® version 9.4 or later will be used for all analysis. Where statistical analyses require the use of R instead of SAS®, R version 4.0.1 or later will be used.

The analyses and outputs specified in this TSAP will be included in the clinical trial report (CTR) unless specified otherwise.

4. CHANGES IN THE PLANNED ANALYSIS OF THE STUDY

[redacted]

The CTP also defined a Full Analysis Set (FAS) for efficacy analyses, which was defined in exactly the same way in the CTP as the Treated Set (TS) for safety analyses, i.e. both were defined as all randomized and treated patients. Reference to only the TS is retained in this document; the TS will be used for all efficacy and safety analyses.

[redacted]

5. ENDPOINTS

5.1 PRIMARY ENDPOINT(S)

The primary endpoint is the improvement (yes/no) from baseline in histological findings based on liver biopsy after 48 weeks of treatment. In practice, patients who discontinued treatment early but after 40 weeks of treatment are expected to have a post-baseline liver biopsy performed, and will also contribute to the primary endpoint derivation.

Patients are expected to have NAS ≥ 4 and a fibrosis stage of F1a, F1b, F1c, F2 or F3 at baseline, according to the entry criteria for the trial.

Improvement in histological findings is defined as a composite of:

1. Improvement in NASH:
    - Decrease of at least 2 points in NAS, with at least 1 point decrease in NAS sub-score of either lobular inflammation or ballooning,

and

2. No worsening of fibrosis
    - Absence of any increase in fibrosis stage.

The NAS and fibrosis scoring system is summarized below. Details of the meaning of each grade/stage are referred to the CTP.

Table 5.1: 1 NAS scoring system for liver biopsies

	Steatosis	**Lobular Inflammation**	**Ballooning**	**NAS (Total Score)**
	Grade 0	Grade 0	Grade 0	
	Grade 1	Grade 1	Grade 1	
	Grade 2	Grade 2	Grade 2	
	Grade 3	Grade 3		
Range	0-3	0-3	0-2	0-8

The total score for NAS is the sum of the 3 NAS sub-scores for steatosis, lobular inflammation and ballooning respectively.

Table 5.1: 2 Fibrosis staging for liver biopsies

	Fibrosis stage	**Fibrosis score**
	Stage F0	0
	Stage F1a	1
	Stage F1b	1
	Stage F1c	1
	Stage F2	2
	Stage F3	3
	Stage F4	4
Range		0-4

The fibrosis score is the same as the assigned stage of fibrosis without the distinction whether a stage of F1a, F1b or F1c was recorded (for a fibrosis score of 1). A patient who has a baseline fibrosis stage of F1a, F1b or F1c and a different post-baseline fibrosis stage of F1a, F1b or F1c is not considered to have worsening or improvement in fibrosis.

Central pathology readings of biopsy slides will be used. These will be physical central pathology readings for patients at sites in all countries, except in China where digital central pathology readings will be used for baseline in the primary endpoint definition (see further below).

All central pathology readings on the database which are considered adequate will be used, including any readings which may be performed by the back-up reader instead of the primary reader. An adequate liver biopsy reading means that the central pathology reader is able to fully read the slides provided (2 different types of staining; 1 per slide), and consequently provide values for all of the 3 NAS sub-scores and for the fibrosis stage, using the CTP guidance on scoring/staging. An inadequate liver biopsy reading will be completely re-read using at least one new stained slide, even if it could be partially read at the first attempt. Consequently, only adequate readings are expected to contribute data for analysis, and missing data at the level of individual items is not expected. In practice, any data records which are not labelled as inadequate on the database will be assumed to be adequate.

For baseline biopsy, either a biopsy performed during the screening period or a historical biopsy (within 6 months) will be used to provide baseline data, without any further distinction made for analysis purposes between these two situations.

For patients at sites in China, a single physical central pathology reading of the screening biopsy is expected to be available as part of the batch reading (screening reading 2). This will

not be used for baseline, since it may generate baseline values which are outside the ranges defined by the entry criteria, which may in turn create further issues with applying endpoint derivations and statistical models. Therefore, for patients at sites in China, the digital central pathology reading of the screening biopsy (screening reading 1) will be used as baseline. The potential limitations of using a digital baseline reading versus a physical post-baseline reading for primary endpoint purposes in China are acknowledged here. For patients at sites outside China, two physical central pathology readings of the same screening biopsy are expected to be available. The first of these (i.e. as used to assess eligibility) will be used for baseline (screening reading 1).

All available post-baseline biopsy data from physical central pathology reading will be used in the primary endpoint definition, with no restriction on how long the biopsy was done after the end of treatment, and using all available post-baseline biopsy results from patients who discontinued treatment early (including any results done earlier than Week 40, in the unlikely event that this situation exists). Patients without a post-baseline biopsy will be imputed as non-responders, regardless of whether a biopsy result was expected or not.

[redacted]

5.2 SECONDARY ENDPOINT(S)

5.2.1 Key secondary endpoint(s)

There are no key secondary endpoints defined in this trial.

5.2.2 Secondary endpoint(s)

Secondary endpoints are defined as follows:

MRI-PDFF

- Improvement in liver fat content (yes/no), defined as a percentage reduction of at least 30% from baseline in liver fat content, measured in units of % using MRI-PDFF, after 48 weeks of treatment.

Liver fat content (mean PDFF) in % is defined as the simple (unweighted) arithmetic mean of values from each of 9 liver segments obtained from a technically acceptable image reading. The mean will be calculated by the vendor and provided in the data transfer. If fewer than 9 liver segments can be read, the mean value will still be calculated, provided that 3 or more liver segments can be read. If fewer than 3 liver segments can be read, the mean will not be calculated and liver fat content (mean PDFF) is then expected to be missing. All non-missing values of liver fat content will be used for analysis, including those values derived from 3-8 liver segments.

Patients with MRI-PDFF data within the Week 48 visit window (see Section 6.7) who meet the above criterion are defined as responders, with no restriction on how long the MRI was done after the end of treatment. and non-responders otherwise. Patients without MRI-PDFF data within the Week 48 visit window will be imputed as non-responders in the main analysis of this endpoint [redacted]

- Absolute change from baseline in liver fat content, measured in units of % using MRI-PDFF, after 48 weeks of treatment.
- Percentage change from baseline in liver fat content, measured in units of % using MRI-PDFF, after 48 weeks of treatment.

Liver biopsy

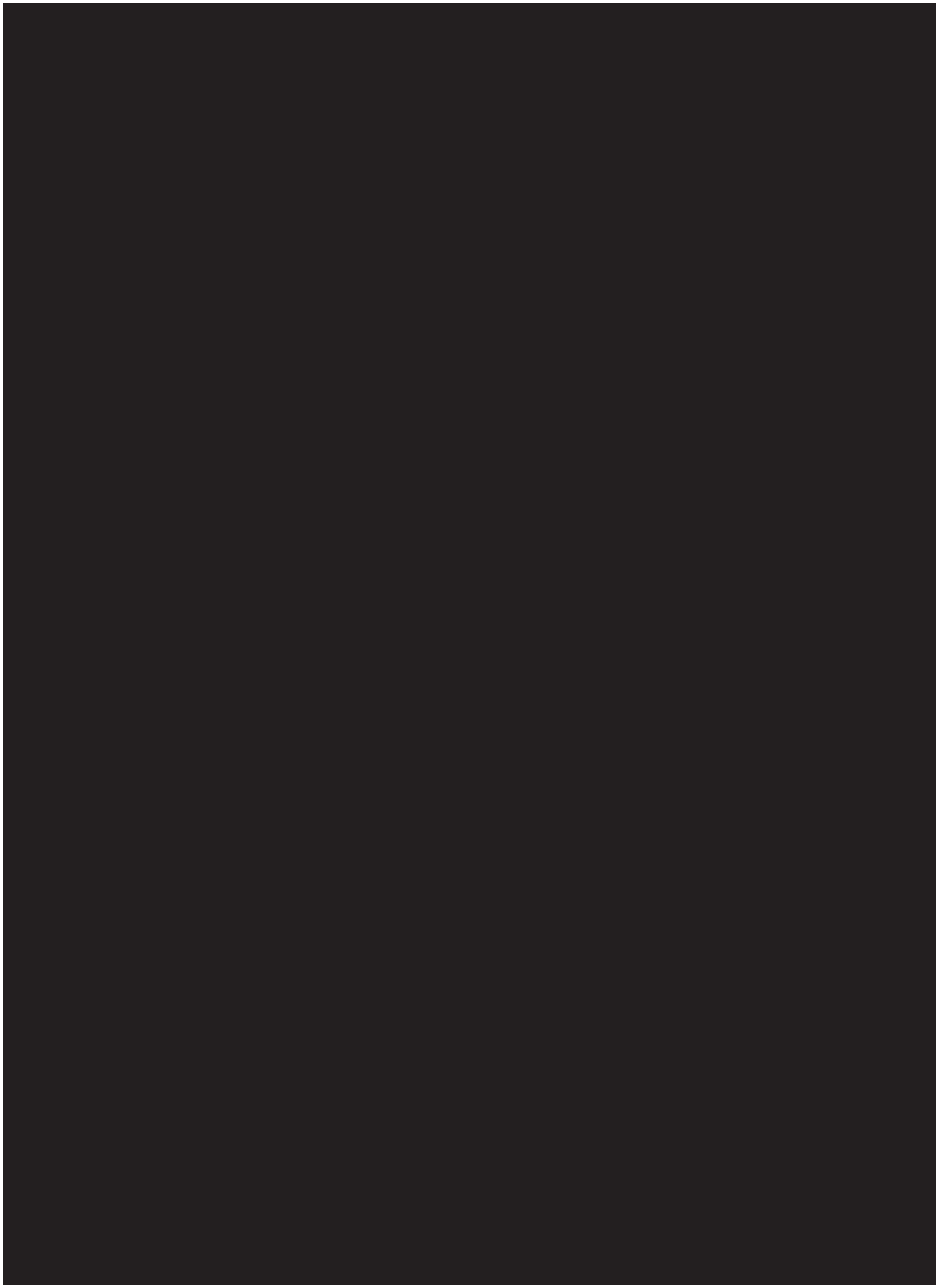
- Improvement in fibrosis (yes/no), defined as at least one stage decrease in fibrosis stage from baseline, assessed by liver biopsy, after 48 weeks of treatment.
- Absolute change from baseline in total score for NAS, assessed by liver biopsy, after 48 weeks of treatment.

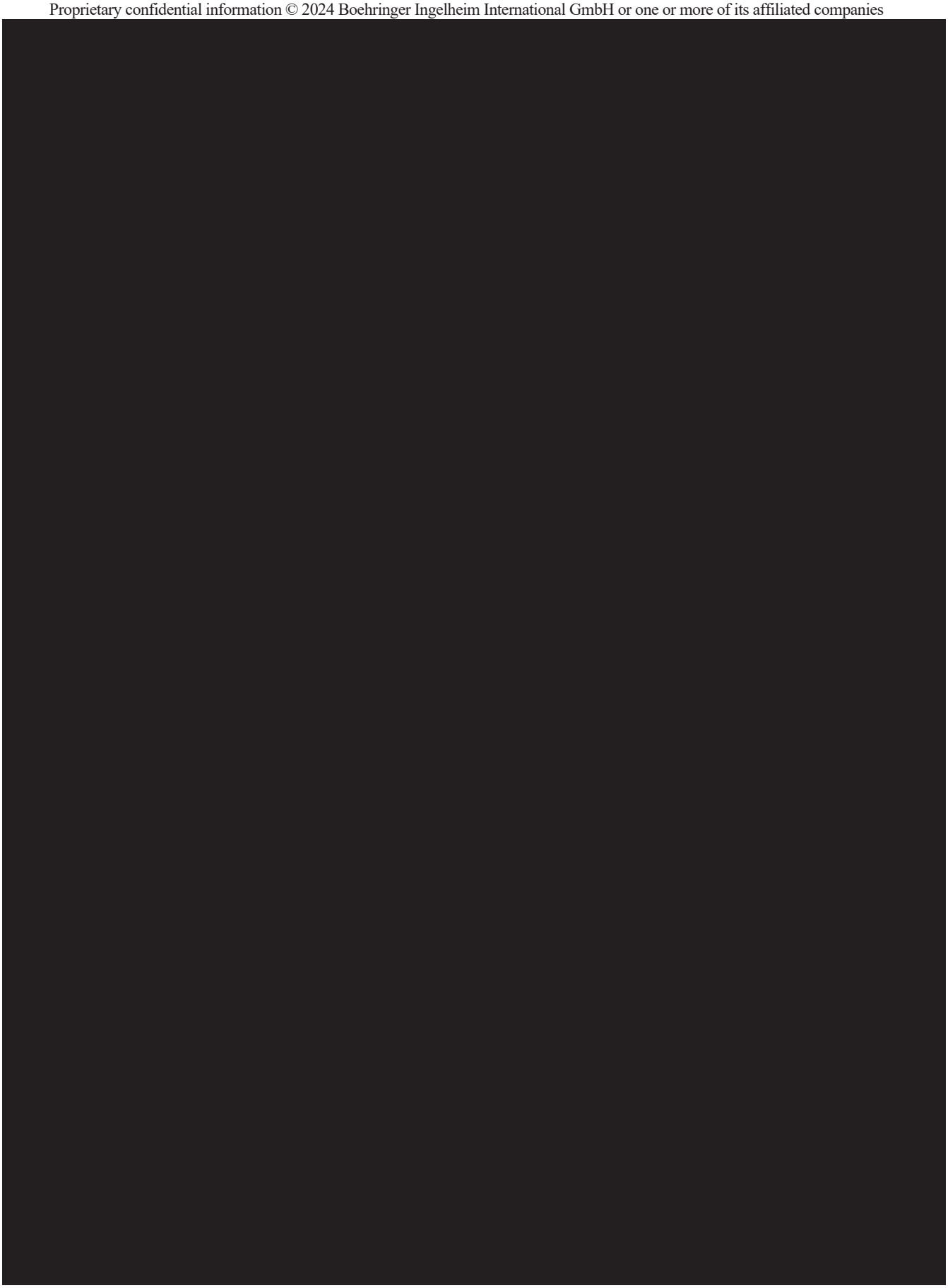
Central pathology biopsy data selection and derivation will be similar to what is described for the primary endpoint in Section 5.1.

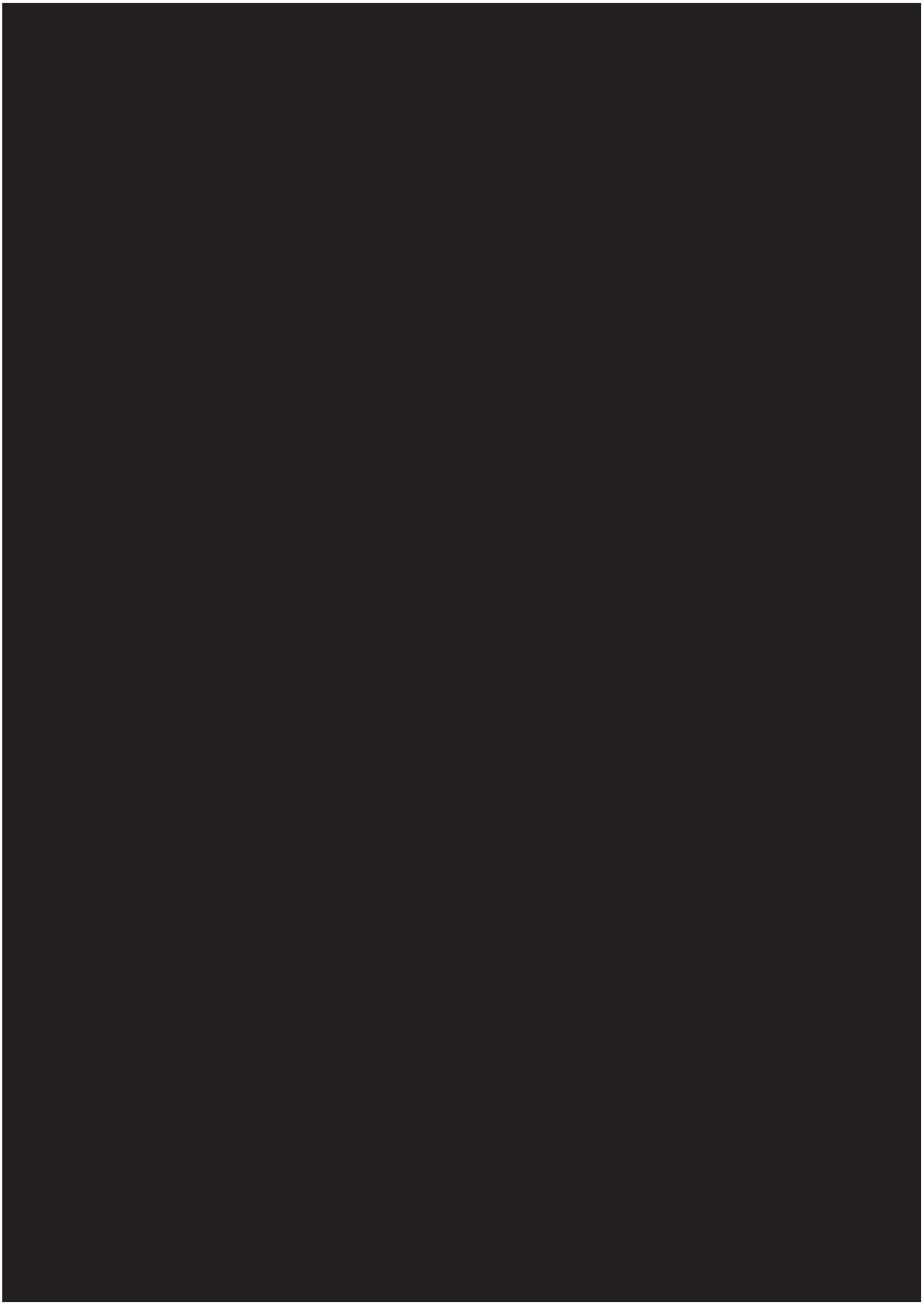
All available post-baseline biopsy data will be used in the secondary endpoint definitions with no restriction on how long the biopsy was done after the end of treatment, and using all available post-baseline biopsy results from patients who discontinued treatment early (including any results done earlier than Week 40).

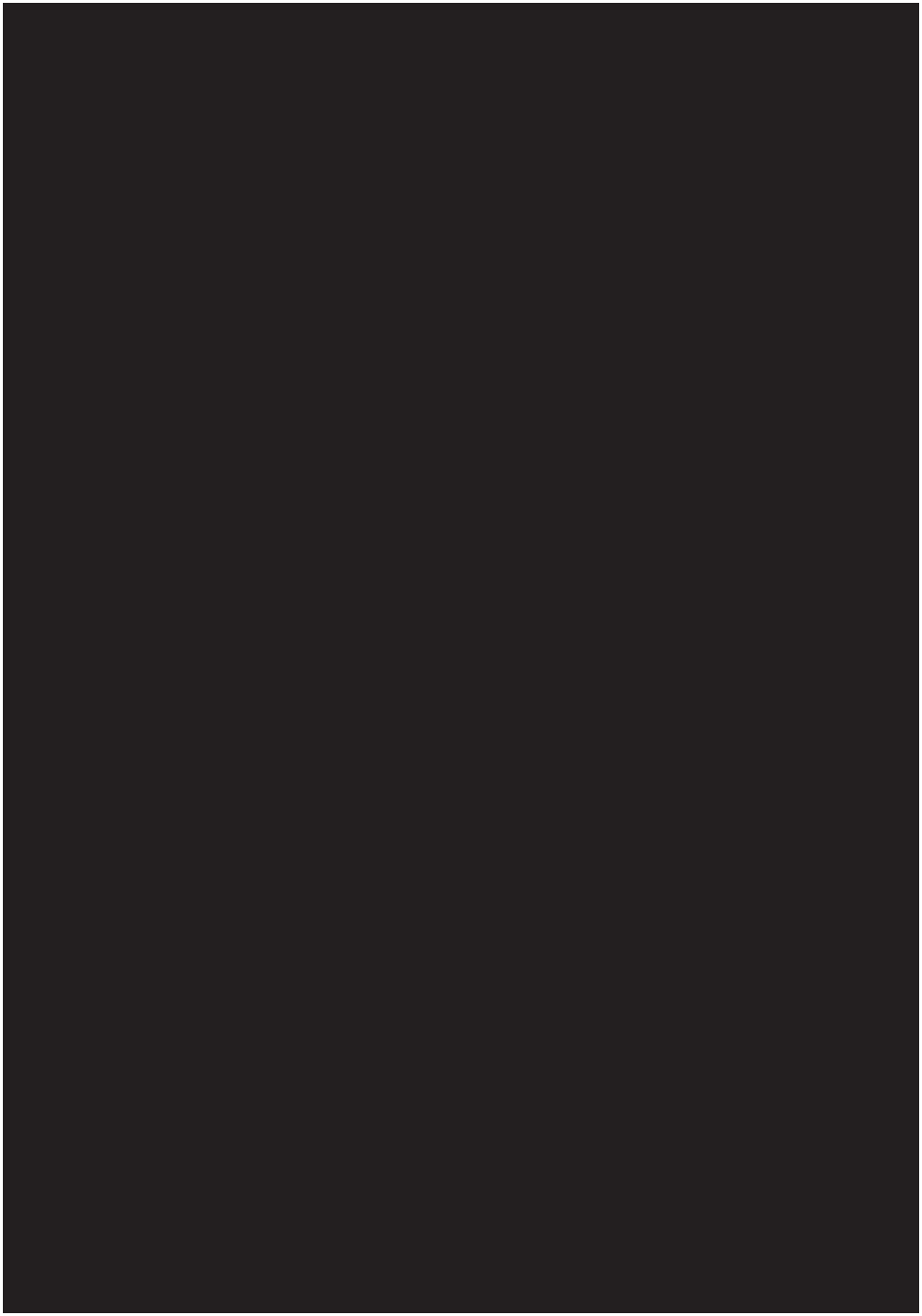
For the responder endpoint, patients without a post-baseline biopsy will be imputed as non-responders, regardless of whether a biopsy result was expected or not.

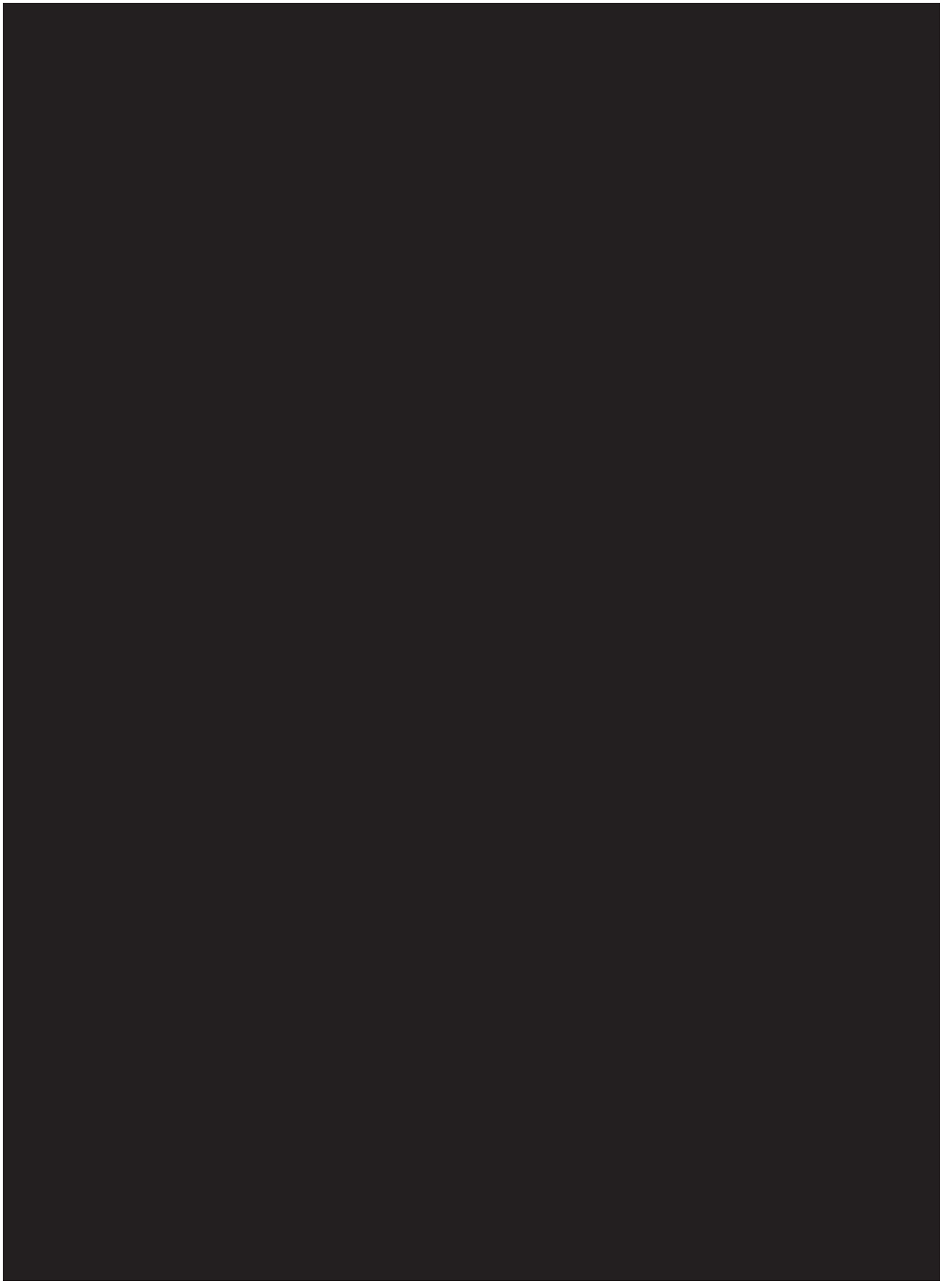
[redacted]











5.4 OTHER VARIABLE(S)

5.4.1 Safety

Safety will be assessed based on:

- Adverse events
  - Routine reporting using the Medical Dictionary for Regulatory Agencies (MedDRA) coding dictionary (version in force at final DBL)
  - Adverse events of special interest (AESIs)
  - Adjudicated adverse events

- Laboratory variables
  - Specific laboratory tests are listed in CTP Table 5.2.3: 1

- Vital signs
  - Pulse rate
  - Systolic blood pressure
  - Diastolic blood pressure

- 12-lead ECG

  Further details regarding 12-lead ECG derivations are given below.

- Columbia - Suicide Severity Rating Scale (C-SSRS)

<u>Laboratory data</u>

Estimated glomerular filtration rate (eGFR) and urine albumin/creatinine ratio (UACR) will be used as provided by the central laboratory, i.e. it will not be re-derived for analysis.

Fib-4 index and APRI will also be used as provided by the central laboratory.

12-lead ECG

Single digital ECG recordings are planned to be collected at each relevant visit with the exception of screening, where triplicate ECG recordings are planned. Each of the recorded single ECGs will then be evaluated for cardiac intervals, which comprise the RR, PR and QT interval and the QRS duration. Measurements of these intervals will be made on four cardiac cycles from the lead chosen (details not specified here).

The measurements of the cardiac cycles will be stored in the database, i.e. four values per visit (for a single ECG). The four cardiac cycles (or however many available) will be averaged, to give a single mean value for each visit. This will be done prior to the calculation below of the heart rate (HR) and heart rate corrected QT interval (QTc).

For screening only, where a triplicate ECG is expected, the mean values obtained for each single ECG will be further averaged across the 3 ECGs (or however many available) to give a single mean value for each variable at screening.

The heart rate will be derived from the RR interval as:

$$\mathrm{HR\ [beats/min]} = \frac{60\,000}{\overline{\mathrm{RR}}}$$

where $\overline{RR}$ is the mean of the four RR intervals (measured in msec).

The QT interval corrected for HR according to Fridericia's formula (QTcF) for a single ECG will be derived as:

$$\overline{\mathrm{QTcF}}\ [\mathrm{msec}] = \left(\frac{1000}{\overline{\mathrm{RR}}}\right)^{1/3} * \overline{\mathrm{QT}}[\mathrm{msec}],$$

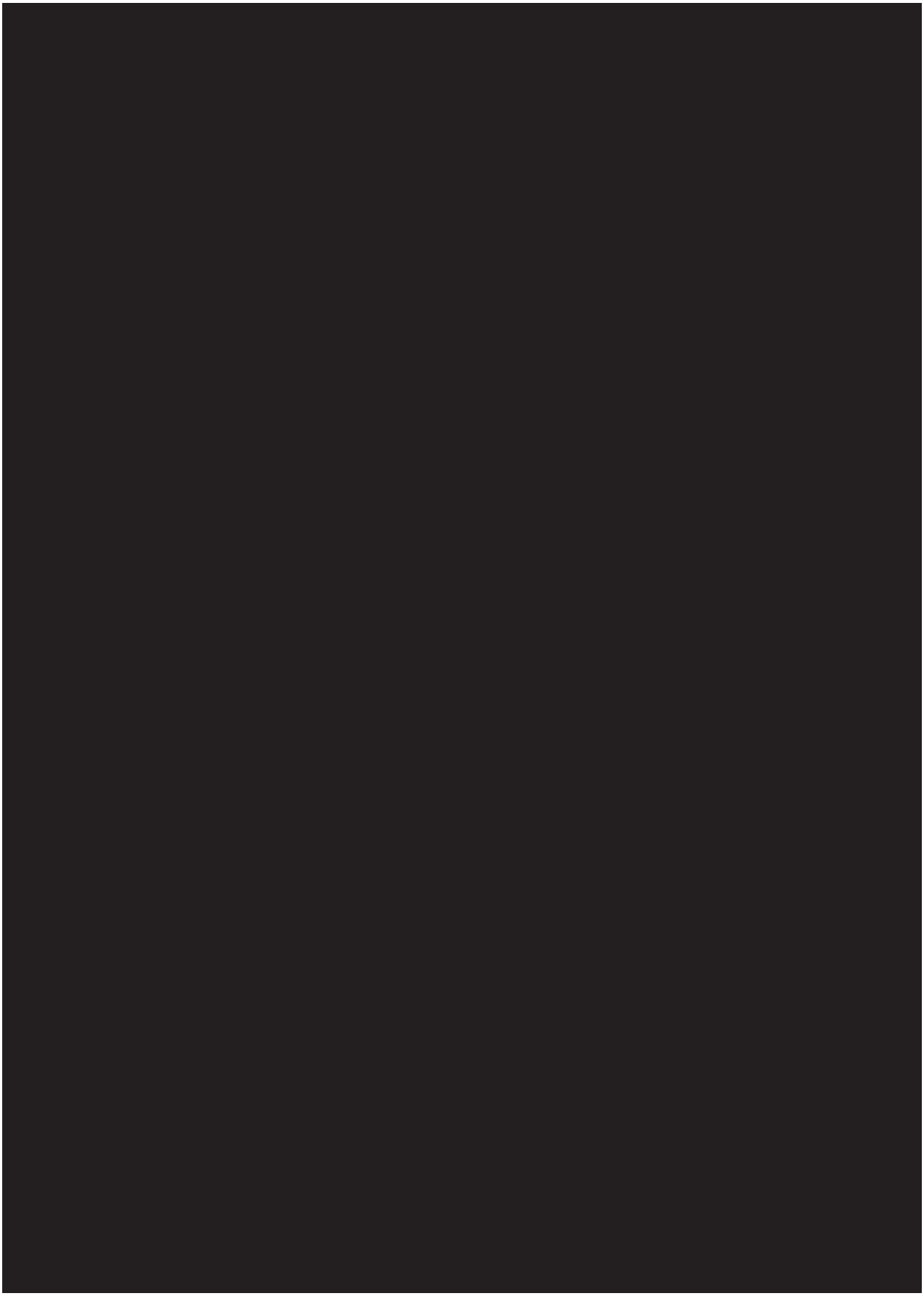
where $\overline{\mathrm{QT}}$ is the mean of the four QT intervals and $\overline{\mathrm{RR}}$ is the mean of the corresponding preceding RR intervals of the four cardiac cycles for this ECG.

Likewise, the HR-corrected QT interval according to Bazett's formula (QTcB) for a single ECG is given by

$$\overline{\mathrm{QTcB}}\ [\mathrm{msec}] = \left(\frac{1000}{\overline{\mathrm{RR}}}\right)^{1/2} * \overline{\mathrm{QT}}[\mathrm{msec}].$$

In the case that more than one ECG is available (i.e. more than one group of cardiac cycles) at a visit where a single ECG was planned, then the (first) ECG on that day which is not labelled as "UNSCHEDULED" will be used, regardless of the reason for this.

New onsets for categorical endpoints are derived based on the tables below:



5.4.2 Demographic and other baseline characteristics

Standard demographic variables and baseline characteristics are used as recorded on the eCRF.

The NAS and fibrosis scoring system is summarised in Section 5.1.

Baseline definitions are given in Section 6.7.

Time since NASH diagnosis will be calculated relative to randomisation date, not relative to informed consent date, since some investigators may only have considered the diagnosis to be confirmed after the screening biopsy was taken.

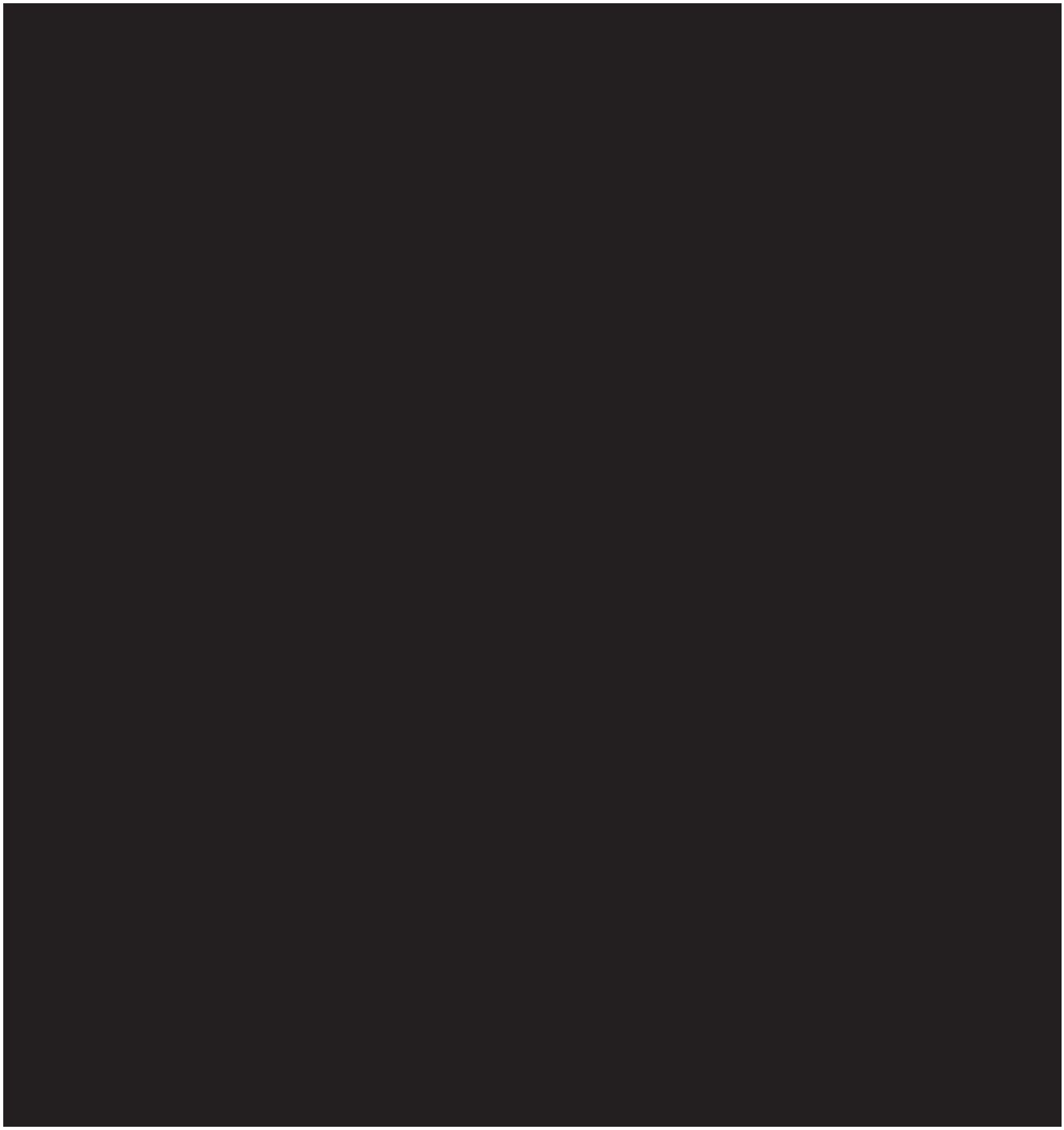
5.4.3 Treatment exposure and compliance

Treatment exposure is defined as the number of days from the first dose of study treatment to the last dose of study treatment inclusive. Any intermediate gaps in treatment in which the patient did not take medication will not be considered. It is assumed that an End of Treatment eCRF page will be available and complete for all treated patients at DBL, and that no special rules are needed to handle any missing information here.

Treatment exposure duration will be categorized as follows:
<4 weeks; 4-<8 weeks; 8-<12 weeks; 12-<16 weeks; 16-<20 weeks; 20-<24 weeks, 24-<28 weeks; 28-<32 weeks; 32-<36 weeks; 36-<40 weeks; 40-<44 weeks; 44-<48 weeks; ≥48 weeks.

Study duration (regardless of whether on-treatment) will also use the same category definitions, although significant off-treatment follow-up after stopping treatment is not expected in this trial.

No treatment compliance calculations will be performed for analysis purposes based on return of unused and empty syringes. Any discrepancies between medication assigned, medication dispensed and medication administered will be assessed through the process for determining IPDs.



5.4.5 Immunogenicity variables

Immunogenicity will be evaluated using anti-drug antibodies (ADA) and neutralising antibodies (NAb). Only ADA is expected to be available for the CTR.

6. GENERAL ANALYSIS DEFINITIONS

6.1 TREATMENT(S)

6.1.1 Study Phases

The following summarises the study phases defined in the protocol for a patient who completes all visits and remains on treatment as planned:

Table 6.1.1: 1 Study phases

Study Phase Description	Start	End
Screening phase	Date of informed consent	Day before Visit 2 (Week 0)
Treatment phase – dose escalation	Date of Visit 2 (Week 0)*	Day before Visit 26 (Week 24)
Treatment phase – maintenance	Date of Visit 26 (Week 24)	Date of Visit 32 (Week 48)
Follow-up phase	Day after Visit 32 (Week 48)	Date of end of study participation (expected to be 4 weeks after Visit 32 (Week 48))

* See also baseline definitions in Section 6.7.

Patients who discontinue treatment early are not required or expected to complete the planned treatment phase. The follow-up phase is defined in the protocol as up to 4 weeks after the end of the planned treatment phase for patients who completed treatment, or 4 weeks after the End of Treatment visit for patients who discontinued early. It is not therefore expected that there will be a significant amount of data captured in this trial for any patients which is more than about a month after stopping treatment.

The follow-up period will not exist for any randomised patient who was lost to follow-up whilst still receiving treatment.

The definition of on-treatment for analysis purposes is given in Section 6.1.6. The use of a residual effect period (REP) means that data captured during the follow-up period specified in the protocol could be assigned to the treatment period for analysis.

6.1.2 Planned Maintenance Treatment Group

Patients are randomised to one of four treatment groups (Placebo, Survodutide 2.4 mg, Survodutide 4.8 mg, Survodutide 6.0 mg). These are the planned maintenance treatment groups (weekly doses).

For the purposes of display in outputs, “planned treatment” means the same as “planned maintenance treatment”.

6.1.3 Actual Maintenance Treatment Group

According to the rules for escalating and adjusting the Survodutide dose given in the CTP, it is envisaged that some patients may not receive the planned maintenance treatment according to the randomised treatment group, whilst nevertheless remaining on treatment.

Actual maintenance treatment group will be defined for analysis as follows:

- If the patient started the maintenance period, then actual maintenance treatment group will be defined in all circumstances to be the dose at the start of the maintenance period (which is expected to be one of the Survodutide maintenance doses), regardless of whether previous dose escalation was performed in the expected way according to the protocol. This includes each of the following situations:
    - Any patient who successfully reached their planned maintenance dose at the end of the dose escalation period
    - Any patient who did not reach their planned maintenance dose at the end of the dose escalation period, and who only reached a lower maintenance dose
    - Any patient for whom the start and end maintenance dose are not the same, regardless of whether the patient started the maintenance period on their planned maintenance dose, and regardless of whether the permanent dose reduction was planned according to the protocol, or whether it was unplanned (e.g. where the wrong tolerability information was given to the IRT and the dose was not re-escalated back to correct the error).

- If the patient discontinued treatment during the dose escalation period, actual maintenance treatment group will be defined as the next planned maintenance dose up from the dose at discontinuation. The same rule will be applied irrespective of the reason for discontinuing treatment during the dose escalation period, and will be applied irrespective of whether the dose prior to the dose at discontinuation was as anticipated from the protocol dose escalation strategy.

Table 6.1.3: 1 Definition of actual maintenance treatment group for patients randomised to active Survodutide maintenance treatment who discontinued treatment during the dose escalation period

Dose at discontinuation	Actual maintenance treatment group
Dose = 0 mg	Placebo
0.3 mg ≤ Dose ≤ 2.4 mg	2.4 mg
2.4 mg < Dose ≤ 4.8 mg	4.8 mg
Dose > 4.8 mg	6.0 mg

- All patients randomised to placebo maintenance treatment or who were switched to placebo during the dose escalation period will be assigned to placebo for actual maintenance treatment group.

For the purposes of display in outputs, “actual treatment” means the same as “actual maintenance treatment”.

The above rules cover all known and anticipated situations. The possibility of other situations which are not anticipated from the protocol, and which cannot be identified from blinded data, cannot however entirely be excluded. A review of unblinded dosing data will be performed immediately after unblinding the database at DBL, to identify any dosing situation for which the above TSAP rules might not not be adequate. Such situations might include, but not necessarily be limited to, unexpected increases in maintenance dose, or starting Survodutide maintenance doses which are not one of 2.4 mg, 4.8 mg or 6.0 mg. An attempt will be made to find out whether data error or omission is the explanation, or whether a real issue actually occurred in any such patients. Any additional decisions required on assigning actual maintenance treatment at that time will be documented in the CTR.

[redacted]

6.1.4 Handling of Treatment Errors and Discrepancies

In addition to the rules in Section 6.1.3 for handling situations envisaged by the CTP, it is also possible that dosing errors may occur. These could either be errors in the site dispensing an incorrect treatment and/or giving incorrect instructions about how to take the dispensed treatment, or they could be patient administration errors.

Dosing errors (other than through incorrect IRT transactions, as described in Section 6.1.3) will not change the assignment of any patient to actual maintenance treatment as described above, regardless of the reason for the error, and regardless of the unblinded dose taken or dispensed at the relevant week(s). However, any such errors and their unblinded doses will be identified as part of the analysis.

Blinded identification of potential discrepancies between assigned, dispensed and administered medication numbers will be performed prior to DBL as part of ongoing monitoring for possible IPDs.

6.1.5 Pooling Across Doses

In addition to specific Survodutide dose columns on descriptive summary tables, a combined column across doses (i.e. “All Survodutide”) will also be presented.

Combined Survodutide dose will not be used in any statistical models or descriptive summaries of efficacy.

Combined Survodutide dose will be included on all disposition and demographic/baseline summary tables, and in descriptive summaries of treatment exposure and safety.

Where a short treatment label is required on outputs, “Survo X.X mg” and “All Survo” will be used for the specific dose groups, and the combination across dose groups, respectively.

6.1.6 Definition of On-Treatment

There is no continued study participation for patients who have discontinued treatment early. Consequently off-treatment data capture is expected to be restricted to those data captured during the safety follow-up period specified by the CTP, although there may be occasional unplanned exceptions to this.

Even for patients who completed treatment according to the protocol, data captured during the follow-up period may be relevant for safety.

It is important to be clear whether all available data will be used in a particular summary or analysis, or whether only on-treatment data will be used. The definition of “on-treatment” for this purpose is as described below, and it includes any temporary treatment interruptions which might occur prior to last dose of study treatment. It is assumed that all patients will have a date of last administration of study medication accurately recorded on the End of Treatment eCRF page.

“Off-treatment” refers to data which are captured after baseline but which are not “on-treatment” as defined below.

Safety

An AE start date or other safety assessment date will be considered “on-treatment” if the AE start date or assessment date is between the date of first dose and 28 days after the date of last

dose of study treatment inclusive (the residual effect period (REP) is defined as 28 days in this trial).

The same on-treatment definition will be used for all types of safety data.

Efficacy

All post-baseline efficacy data (including for exploratory biomarkers) will use the same 28-day definition of "on-treatment" as for safety, where on-treatment analyses are specified (as opposed to those efficacy analyses which use all data, both on- and off-treatment).

6.2 IMPORTANT PROTOCOL DEVIATIONS

Handling of IPDs is included in the DV domain specifications and stored within the TMF in EDMS.

IPDs will not lead to exclusion from analysis sets.

6.3 SUBJECT SETS ANALYSED

The following analysis sets will be defined:

Screened Set

The screened set (SS) includes all patients for whom a signed informed consent is available at the time of final DBL. It will be used only for display of patient disposition.

Randomised Set

The randomised set (RS) includes all screened patients who were randomised in the trial, regardless of whether any study drug was taken. It will be used only for display of patient disposition and IPDs.

Treatment assignment will be as randomised (i.e. planned maintenance treatment).

Treated Set

The treated set (TS) is defined as all randomised patients who received at least one dose of study treatment.

The TS will be used for presentation of demographic/baseline data, concomitant medications/diagnoses, treatment exposure, efficacy and safety.

Treatment assignment will be according to both actual maintenance treatment and planned maintenance treatment (i.e. as randomised), depending on the analysis.

It is not expected on the database that the situation in which a patient could be treated without randomisation will exist. Documentation is in place by the Interactive Response Technology (IRT) vendor, which states that, should any patient be given treatment at the site in error

without randomisation, such a patient will be subsequently randomised to a treatment which is not incompatible with the one given in error.

Full Analysis Set

The full analysis set (FAS) was defined in the CTP in the same way as TS. Since the definitions are the same, the TS label will be retained for both efficacy and safety analysis, and the FAS label will not be used.

Per Protocol Set

A per protocol set (PPS) will not be defined in this study.

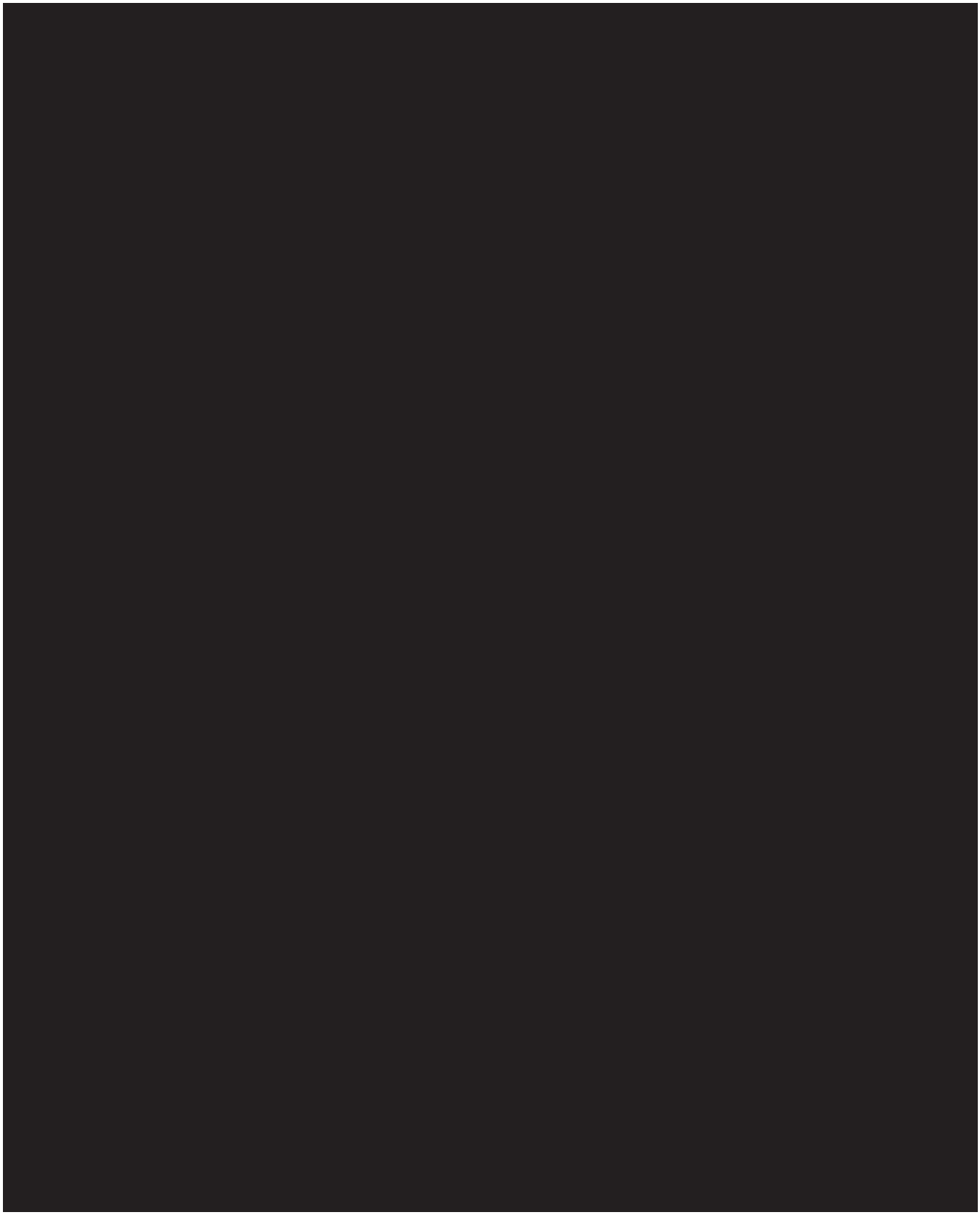
[redacted]

[redacted]

The table below summarises the analysis sets and their main intended purpose [redacted]

Table 6.3: 1 Subject sets analysed

Data type	**Analysis set**			
	SS	RS	TS	[redacted]
Disposition	x	x		
Demographic and baseline characteristics			x	
Concomitant diseases and medications			x	
Exposure to study drug			x	
Primary efficacy endpoint			x	
Secondary efficacy endpoints			x	
[redacted]			[redacted]	
[redacted]			[redacted]	
Adverse events			x	
Vital signs			x	
Laboratory safety data			x	
12-lead ECG			x	
Other safety			x	
[redacted]				[redacted]



6.5 POOLING OF CENTRES

No analysis will be performed in which centre is included in the statistical model.

Subgroup analysis will be performed by country and region, but no other pooling of centres will be performed.

6.6 HANDLING OF MISSING DATA AND OUTLIERS

Efficacy

For responder efficacy endpoints using biopsy, MRI-PDFF and [redacted] (including the primary endpoint defined in Section 5.1), patients without data to evaluate the response will be imputed as non-responders (whether the missing data are expected from the CTP schedule or not). For MRI-PDFF this simple non-responder imputation will be used in the main analysis. [redacted].

For continuous efficacy endpoints, missing data will be handled implicitly by the statistical model (mixed model for repeated measures (MMRM)). This model makes the assumption that the data are MAR. Patients who do not have a value at the time point of interest, but who have earlier post-baseline values, still contribute to this analysis. MAR means that a patient who drops out is assumed to have behaved like a similar patient within the same treatment group (where "similar" means both with regard to baseline covariates in the model and to the observed trajectory of post-baseline values prior to drop out).

For continuous endpoints derived from MRI-PDFF and [redacted] (only), [redacted]

No other imputations will be performed for any efficacy endpoint.

Safety

A completely missing concomitant medication start date will be set to screening date. If only the month and year are available for the AE/concomitant medication start date, the first day of the month will be assumed. If only the year is available, then 1st January of that year will be assumed. In the case of AE start dates (only), if the stated imputation would make the AE become pre-treatment when this was not known from the incomplete date, it will be assumed instead that the AE started on the first day of treatment.

If only the month and year are available for date of last administration of study medication, the last day of the month will be assumed.

If only the month and year are available for the date of first administration of study medication, the date of randomisation will be used if they are in the same month. If randomisation was in the month prior to first drug administration, the first day of the month of first administration will be assumed.

[redacted]

6.7 BASELINE, TIME WINDOWS AND CALCULATED VISITS

Definition of baseline

In general, baseline is defined as the last non-missing measurement on or prior to the first administration of study medication. In the case that the first dose of study medication is delayed beyond the date of randomisation, a post-randomisation value can be used as a baseline value provided that it meets the stated requirement.

Where clock times are available (expected to include laboratory safety data, 12-lead ECG, plasma Survodutide concentrations and biomarkers) they will be used in addition to dates to define baseline, since it is also expected that the time of first study drug administration will be accurately recorded on the eCRF for all patients.

Unscheduled and repeat assessments are eligible for use as baseline values provided that they meet the above criterion.

For 12-lead ECG, if a screening value needs to be used for baseline, it will be the average of all available triplicate screening ECG measurements (see Section 5.4.1 for derivation details).

For vital signs, the baseline definition will make the additional requirement that the data records are pre-dose. In practice, this will also be done using the clock times as well as the date. A single pre-dose measurement is expected to be captured at the baseline visit (Visit 2), which will therefore be used as baseline. In any situation in which only a post-dose measurement is available at Visit 2 and no pre-dose measurements, Visit 2 will not be used as baseline; the screening visit (Visit 1) will be used as baseline instead.

For baseline NAS and fibrosis scores, the following specific considerations will be applied:

- Either a biopsy performed during the screening period or a historical biopsy (within 6 months) will be used to provide baseline data, without any further distinction made for analysis purposes between these two situations (both referred to subsequently as “screening biopsy”).

- The baseline value will be derived from central pathology readings of screening biopsy slides. For sites in all countries except China, central pathology screening readings will always be physical readings. For sites in China, central pathology screening readings may be digital readings, the handling of which is described below.

- All adequate central pathology readings which meet the baseline definition will be used, including any (expected to be occasional) readings performed by the back-up

reader instead of the primary reader. The meaning of "adequate" is given in Section 5.1.

- The first screening central pathology reading (i.e. as used to assess eligibility) will be used for baseline in all cases where a re-reading of the screening biopsy was subsequently done, unless otherwise specified for any analysis. For sites in China (only), the first screening central pathology reading is a digital reading and this will be used.

MRI-PDFF and [redacted] procedures are only scheduled to be performed at screening, and not also at the randomisation visit. Therefore, the baseline value is the screening value in these cases.

Visit windows

All summaries and analysis which are presented by week will use a visit window to classify the data record. This approach allows appropriate classification of visits which may have occurred significantly earlier or later than the protocol assessment schedule, as well as the use of data captured at visits which have no fixed timing (notably the EOT visit). Relative day will be defined for this purpose as *(date of assessment – date of first administration of study medication) + 1*.

Nominal database visit labels will not be used in any summary or analysis by time point, unless otherwise stated.

Prior to applying the rules for handling multiple values in visit windows specified below, it will first be determined whether the data record is on-treatment, as defined in Section 6.1.6, for all summaries or analysis which require this approach.

Unscheduled and repeat assessments (where relevant/available) are eligible for inclusion in visit windows.

For all variables except 12-lead ECG, if a patient has more than one non-missing value within the same visit window, the value closest to the target day will be selected. If two non-missing values are the same distance from the target day, the earlier of the two values will be selected.

For 12-lead ECG (only), scheduled visits take precedence over unscheduled visits. Once this rule has been applied, the measurement closest to the target day will be used. If there are two observations which have the same difference in days to the planned day (but are not on the same day), the later of the two values will be selected.

Derivation of last, minimum and maximum values where relevant in safety analyses will consider all available values, regardless of whether they were selected in the visit window.

The derivation of the maximum value of an ECG interval (or worst case of the morphological assessment, if applicable) will consider all on-treatment values (whether or not selected in any time window; see Section 6.1.6 for definition of on-treatment).

Four different situations are envisaged, which are described in the following tables:

- Visit windows for variables scheduled to be assessed at all or most clinic visits
- Visit windows for biomarker variables
- Visit windows for variables with an intermediate clinic visit schedule
- Visit windows for variables with a sparse clinic visit schedule

Table 6.7: 1 Visit windows for variables scheduled to be assessed at all or most visits

Period	Time Point	Target Day	Visit Window
Dose escalation	Week 1	8	2-11
	Week 2	15	12-18
	Week 3	22	19-25
	Week 4	29	26-35
	Week 6	43	36-49
	Week 8	57	50-63
	Week 10	71	64-77
	Week 12	85	78-98
	Week 16	113	99-126
	Week 20	141	127-154
Maintenance	Week 24	169	155-182
	Week 28	197	183-210
	Week 32	225	211-238
	Week 36	253	239-266
	Week 40	281	267-294
	Week 44	309	295-322
	Week 48	337	≥ 323

The relevant variables are expected to include vital signs, routine laboratory safety variables (all expected to be non-fasting), 12-lead ECG and the mental health questionnaire (C-SSRS).

In particular, the Week 2 and Week 4 visit windows from this table will still be used for laboratory safety variables and 12-lead ECG, even though these are not also scheduled at Week 1 and Week 3.

Table 6.7: 3 Visit windows for variables with an intermediate visit schedule.

Period	Time Point	Target Day	Visit Window
Dose escalation	Week 8	57	2-84
	Week 16	113	85-154
Maintenance	Week 28	197	155-266
	Week 48	337	≥ 267

The relevant variables are expected to be body weight / BMI derived from this, waist circumference and hip circumference.

[redacted]

Table 6.7: 4 Visit windows for variables with a sparse visit schedule.

Period	Time Point	Target Day	Visit Window
Maintenance	Week 28	197	2-266
	Week 48	337	≥ 267

The relevant variables are expected to be MRI-PDFF and the [redacted].

The use of such a wide visit window for MRI-PDFF at Week 28 means that it will include and display/analyse as "Week 28" any MRI assessments which were done unexpectedly early.

Post-baseline biopsy

Patients are scheduled to provide data from a single post-baseline biopsy at Week 48, which may be earlier (expected 40 weeks onwards) for patients who discontinued early, but nevertheless remained on treatment for long enough to justify performing the biopsy. Two different ways of handling this in analysis are planned, and these are expected to be identical if the protocol requirements are followed exactly:

- All post-baseline biopsy data are used, regardless of how early the biopsy was done (this is equivalent to specifying a single visit window for "Week 48" as ≥ 2 days)
- Only post-baseline biopsy data are used which are after 40 weeks (this is equivalent to specifying a single visit window for "Week 48" as ≥ 280 days)

Vital signs

For vital signs, the following windowing rules will be applied within each visit to determine pre-dose and 10 minutes post-dose from recorded clock times:

- If the vital signs measurement time is prior to or equal to the time of study drug administration, then it is considered eligible for pre-dose, without restriction on how long before study drug administration the measurement was done (provided it was still done on the same visit date). If this selects more than one value as "pre-dose", the latest will be used.
- If the vital signs measurement time is after the time of study drug administration, then it is considered eligible for post-dose, without restriction on how long after study drug administration the measurement was done (provided it was still done on the same visit date). If this selects more than one value as "post-dose", the value closest in time to the target 10 minute post-dose time point will be used.

The situation in which vital signs measurement time is exactly equal to the time of study drug administration is not clinically plausible, and a rule for handling it is included above only for completeness, in case this cannot be resolved for any patients and this situation remains on the database at final DBL.

Any vital signs record for which no time of study drug administration exists (expected to include measurements recorded during the 28 day REP and scheduled Week 48 measurements) will be considered “pre-dose” according to the above rules. Only pre-dose measurements will be used for efficacy purposes (blood pressure further endpoints).

7. PLANNED ANALYSIS

Descriptive statistics for continuous variables will generally include N (number of patients with non-missing values), arithmetic mean, standard deviation (SD), minimum, Q1 (lower quartile), median, Q3 (upper quartile) and maximum. In general, means, medians, other quartiles and SDs will be presented to one more decimal place than the raw data. Minimums and maximums will be presented to the same number of decimal places as the raw data.

[redacted]

Tabulations of frequencies for categorical data will include all possible non-missing categories unless otherwise specified, and will display the number of patients in a category as well as the percentage (%) relative to the respective treatment group. Percentages will be rounded to one decimal place. A missing category will be displayed only if there are actually missing values. Percentages will be based on all patients in the respective patient set whether they have non-missing values or not.

Disposition

The disposition of patients will be displayed. This table will present the number of patients enrolled, randomised and treated. The number treated will be further presented separating the dose escalation and maintenance periods. Planned maintenance treatment (see Section 6.1.2) will be used for this table.

The number of patients who completed or who prematurely discontinued study treatment will be summarised, along with the categorization of the primary reason for treatment discontinuation. The number of patients and reasons for prematurely discontinuing treatment will also be displayed separately for the dose escalation and maintenance periods.

The number of patients who completed or who prematurely withdrew from the study will be summarised, along with the categorization of the primary reason for study withdrawal. It should be noted that in this study, a patient is considered to have completed the study if all scheduled safety follow-up procedures have been performed for that patient, including for patients who discontinued treatment early. In other words, a patient may complete this study without completing treatment.

Where shown on the disposition table, the percentages will be of the number of patients treated in each treatment group.

Kaplan-Meier estimates will be presented graphically by planned maintenance treatment group for time to early treatment discontinuation (defined as *treatment discontinuation date*

– *treatment start date + 1*). In this analysis, patients who did not discontinue treatment early will be censored using the date on which they completed treatment.

The frequency of patients in each of the different analysis sets (Section 6.3) will also be tabulated by planned maintenance treatment.

A shift table will be produced to display differences between planned maintenance treatment and actual maintenance treatment.

A shift table will also be produced to display differences between the first and last doses of maintenance treatment (for patients who reached the maintenance period). This shift table will be produced separately using both dispensed and administered dosing.

IPDs will be summarised and listed for the randomised set using planned maintenance treatment.

7.1 DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS

Descriptive summaries of the following demographic variables and characteristics measured at baseline will be presented, to include: sex, race and ethnicity, age (years) (continuous and categories), height (cm), weight (kg) (continuous and categories), BMI (kg/m2) (continuous and categories), systolic BP (mmHg), diastolic BP (mmHg), pulse rate (bpm), country and region.

[redacted]

In addition, the overall baseline NAS and 3 NAS sub-scores, the baseline fibrosis stage and the time since NASH diagnosis will be summarised.

Summaries of demographic data and baseline characteristics will be presented on the TS using both planned maintenance treatment and actual maintenance treatment.

7.2 CONCOMITANT DISEASES AND MEDICATION

Only descriptive statistics are planned (number and percentage of patients). Summaries of concomitant diseases (medical history and baseline conditions), concomitant medication and concomitant non-drug therapy will be based on the TS, and will all use both of actual maintenance treatment and planned maintenance treatment.

Baseline conditions and concomitant non-drug therapies will be coded according to the most recent version of MedDRA at the time of final DBL. This will include coding of trial-specific recording of medical history (“Medical History – Specific Conditions” eCRF page) for the relevant summary table (see below).

Concomitant medications will be coded according to the most recent version of the World Health Organisation (WHO) Drug Dictionary at the time of final DBL.

Baseline conditions and medical history will be summarised descriptively by MedDRA system organ class (SOC) and preferred term (PT), which will include coded trial-specific

medical history. In addition, trial-specific medical history will be summarised using reported term from the eCRF page (i.e. Diabetes Mellitus Type 1, Diabetes Mellitus Type 2, Diabetes Mellitus Other or Unknown, Arterial Hypertension, Hyperlipidaemia, Cardiac Dysrhythmias, Metabolic Syndrome).

A medication or non-drug therapy will be considered concomitant to treatment if either of the following applies:

- It is ongoing at the start of study treatment
- The start date occurs between the start of study treatment and 28 days after the end of study treatment inclusive.

A medication or non-drug therapy will be considered prior to treatment if the end date occurs before the start of study treatment.

Concomitant medication will be summarised descriptively by ATC3 class and preferred name. Prior medication will not be summarised.

Concomitant non-drug therapy will be summarised by SOC and PT. Prior non-drug therapy will not be summarised.

7.3 TREATMENT COMPLIANCE

Treatment compliance based on return of unused and empty syringes will not be analysed in this trial.

Any discrepancies between medication assigned, medication dispensed and medication taken will be assessed through the process for determining important protocol deviations (IPDs).

Section 6.1.3 describes situations in which the actual maintenance treatment may not be the same as the planned (randomised) maintenance treatment. The CTP specifies rules for adjusting the Survodutide dose (or if necessary discontinuing the patient from treatment) depending on the number of missed doses and whether missed doses are consecutive. Therefore treatment compliance is directly related to the procedure for assigning Survodutide dose.

7.4 PRIMARY ENDPOINT(S)

Frequency tables will be produced of the primary endpoint (improvement in histological findings based on liver biopsy after 48 weeks of treatment, as defined in Section 5.1) in each of the following different ways:

- TS, actual maintenance treatment, using all post-baseline biopsy data (both on- and off-treatment as defined in Section 6.1.6, and irrespective of timing relative to first dose of study treatment)

- TS, planned maintenance treatment, using all post-baseline biopsy data (as above)

- TS, actual maintenance treatment, using only on-treatment post-baseline biopsy data which are at least 280 days (40 weeks) after first dose of study treatment

- TS, planned maintenance treatment, using only on-treatment post-baseline biopsy data which are at least 280 days (40 weeks) after first dose of study treatment.

The above frequency tables will display the number and proportion of responders and non-responders respectively. Non-responders will be further split on these tables into those derived from the available data without imputation and those which were imputed due to a missing response.

7.4.1 Primary analysis of the primary endpoint(s)

The primary analysis of the primary endpoint will be performed on the TS using actual maintenance treatment, and using all post-baseline biopsy data (as above). Patients without post-baseline biopsy (whether this is expected or not) will be imputed as non-responders in this analysis.

A logistic regression will first be performed on the primary endpoint to estimate the treatment effects of multiple doses of Survodutide and placebo, together with the corresponding variance-covariance matrix. Applying the Multiple Comparison Procedures and Modelling (MCPMod) approach, these estimates will then be further used to test for a non-flat dose response curve and to identify suitable dose-response shapes out of a selection of candidate models.

The MCPMod approach [R10-1424, R15-4293] is implemented in two main steps: (1) trial design stage; (2) trial analysis stage. The procedures for the trial design stage, including the selection of candidate models covering a suitable range of dose-response shapes, are provided in the CTP Section 7.2.2. Additional details for the trial analysis stage are specified below.

Logistic regression

The responses for each dose group, as well as their variance-covariance matrix, for the primary endpoint will be estimated using a logistic regression (using a logit link via PROC LOGISTIC in SAS®). The logistic regression model will be fitted without an intercept, and will include presence of diabetes of any type [yes, no], baseline fibrosis score [F1, F2, F3] and the dose group as factors. In the case that 0 events are observed in any combination of dose group and strata, a penalized regression based on the Firth bias reduction method [R20-2291] will be used. The subsequent MCPMod analysis will be fully carried out on the logit scale, and the graphical displays of the fitted models in the Mod step will then be converted from the logit scale to the probability scale for interpretation reasons.

A summary table will be produced of the logistic regression results, including estimated odds ratios for the comparison of each Survodutide dose against placebo and their 95% confidence intervals and associated p-values.

MCPMod

The “DoseFinding” package [R15-2001] will be used for MCPMod analysis.

The multiple comparison procedure will be implemented using contrast tests which control the family-wise type I error rate at a one-sided $\alpha = 0.05$. Candidate models will be used as described in the CTP Section 7.2.2, that is:

- Linear
- Exponential1: 25% of maximum effect achieved at dose 3.0 mg (note: this dose is used for model specification only, it was not included in the trial)
- Exponential2: 5% of maximum effect achieved at dose 3.0 mg
- Emax1: 50% of maximum effect achieved at dose 3.0 mg
- Emax2: 80% of maximum effect achieved at dose 3.0 mg
- Quadratic: Maximum effect achieved at dose 4.8 mg.

For the MCPMod test, the contrast coefficients of each candidate model will be calculated using the R-function 'optContr' based on observed results from the logistic regression (with Firth bias reduction if used), that is, the variance-covariance matrix from the logistic regression step will be used as input to 'optContr'.

If at least one dose-response model is statistically significant, the null hypothesis of a flat dose-response curve can be rejected indicating a benefit of Survodutide over placebo.

All dose-response models will be re-fitted to the data (irrespective of statistical significance in the previous step) without any parameter assumptions to generate a set of new estimates of the model parameters. For all significant models, the generalised Akaike Information Criterion (AIC) will be calculated and reported in the displays. The final model will be chosen as the model with the best fit according to AIC, that is, the model with the smallest AIC value.

The target dose for MCPMod purposes is defined as the smallest Survodutide dose which is estimated to achieve an absolute improvement in the proportion of patients with a histological improvement based on liver biopsy after 48 weeks of treatment of 20% or more compared with placebo. If no dose within the range studied by the trial meets this criterion, then it will be concluded that MCPMod has not identified a target dose.

Estimates for each dose group will be calculated and will be based on the final dose-response model. In practice, the choice of the target dose to be investigated in Phase 3 will be based upon efficacy as well as considering tolerability and safety and other relevant information.

The following tables and displays are planned:

- Table of the contrast coefficients per dose group and candidate model, together with the MCPMod test statistics and p-values for each model, and the critical value.

- For all models fitted to the observed data: Figure of the fitted dose-response curve plus 95% confidence band overlayed with adjusted mean and corresponding 95% confidence interval per dose (estimated from the logistic regression (with Firth bias reduction if used)) – this figure is plotted on the logit scale.

- For all models fitted to the observed data: Figure of the fitted dose-response curve plus 95% confidence band overlayed with adjusted mean and corresponding 95%

confidence interval per dose (estimated from the logistic regression (with Firth bias reduction if used)) – this figure is plotted on the probability scale (figure converted from logit-scale to probability scale).

The predicted Survodutide dose which achieved a 20% improvement in absolute responder proportions compared with placebo will be annotated onto the above figures on the probability scale for the significant model shapes (Linear, Exponential, Emax and Quadratic).

7.4.2 Sensitivity analysis, subgroup analysis, exploratory analysis of the primary endpoint(s)

Sensitivity analysis

The analysis of the primary endpoint specified in Section 7.4.1 will be repeated for the following, using the same model for estimation (logistic regression with Firth bias reduction if necessary) unless otherwise specified, with the same model terms, and using the same subsequent MCPMod procedure:

- Planned maintenance treatment; using all post-baseline biopsy data (as above)

7.5 SECONDARY ENDPOINT(S)

7.5.1 Key secondary endpoint(s)

This section is not applicable as no key secondary endpoint has been specified in the protocol.

7.5.2 Secondary endpoint(s)

MCPMod dose finding analyses will only be performed for secondary endpoints where specified below. [redacted]

MRI-PDFF responder secondary endpoint - main analysis

Frequency tables will be produced of the MRI-PDFF responder secondary endpoint (percentage reduction of at least 30% from baseline in liver fat content, measured in units of % using MRI-PDFF, after 48 weeks of treatment) in both of the following ways:

- TS, actual maintenance treatment, using all post-baseline MRI-PDFF data within the Week 48 visit window (see Section 6.7), both on- and off-treatment (see Section 6.1.6)
- TS, planned maintenance treatment, using all post-baseline MRI-PDFF data within the Week 48 visit window, both on- and off-treatment.

The above frequency tables will be displayed similarly to those for the primary endpoint.

The frequency tables will be produced overall, and further split by each of the subgroups defined in Section 6.4.

Similarly to the primary endpoint analysis described in Section 7.4.1, a logistic regression will first be performed on the MRI-PDFF responder secondary endpoint to estimate the treatment effects of the Survodutide doses and placebo, together with the corresponding variance-covariance matrix. The MCPMod approach will then use these estimates to test for a non-flat dose response curve and to identify suitable dose-response shapes out of a selection of candidate models.

The logistic regression model will be fitted without an intercept, and will include presence of diabetes of any type [yes, no], baseline fibrosis score [F1, F2, F3] and the dose group as factors, and baseline liver fat content (assessed by MRI-PDFF) as a continuous linear covariate. In the case that 0 events are observed in any combination of dose group and strata, the Firth bias reduction method will again be used. The subsequent MCPMod analysis will again be fully carried out on the logit scale, with conversion from the logit scale to the probability scale for interpretation.

A summary table will be produced of the logistic regression results, including estimated odds ratios for the comparison of each Survodutide dose against placebo and their 95% confidence intervals and associated p-values.

The multiple comparison procedure will again be implemented using contrast tests which control the family-wise type I error rate at a one-sided $\alpha = 0.05$. The same candidate models will be used as those used for the primary endpoint MCPMod analysis (described in Section 7.4.1), and the same method will be used to calculate the contrast coefficients.

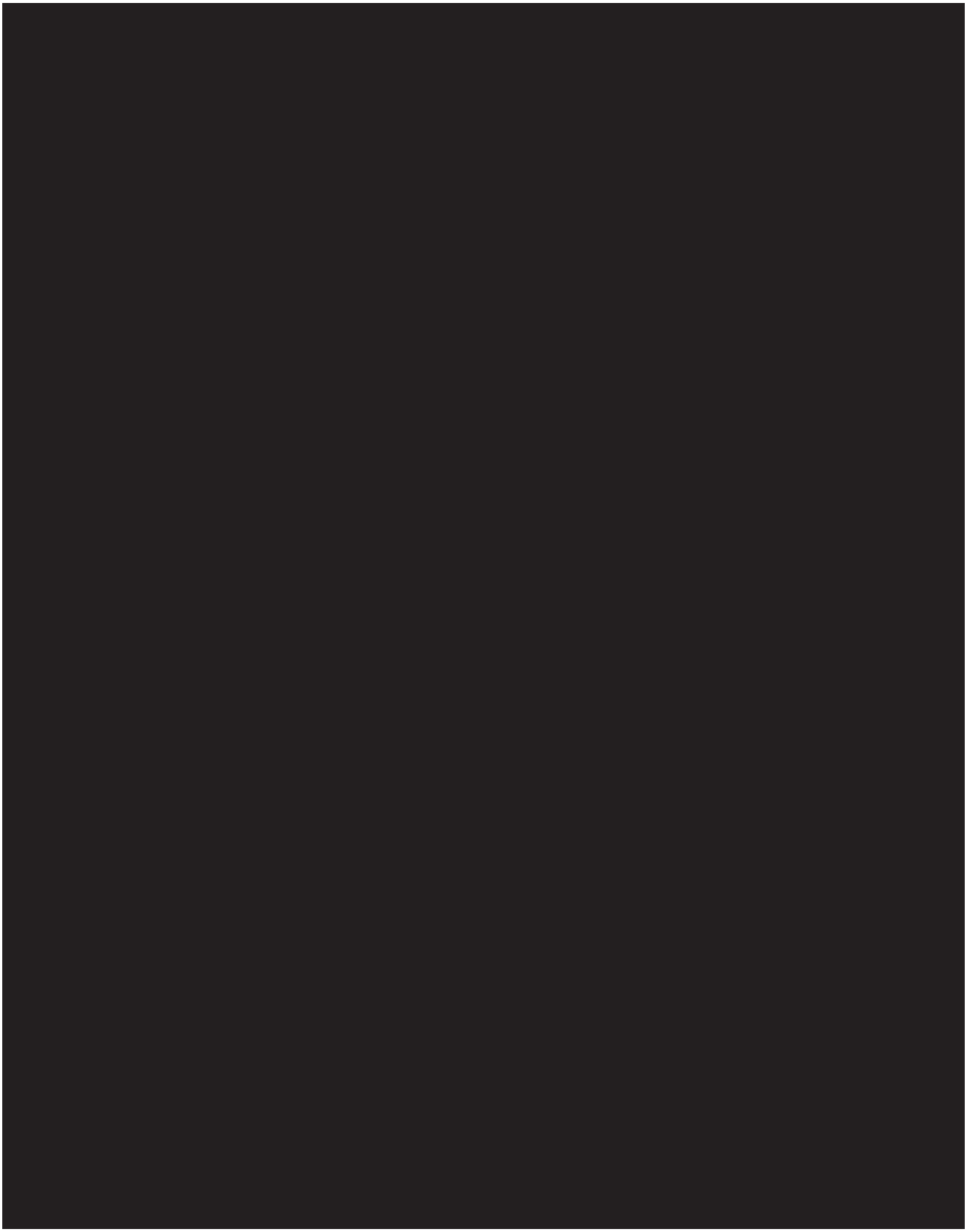
If at least one dose-response model is statistically significant, the null hypothesis of a flat dose-response curve can be rejected indicating a benefit of Survodutide over placebo.

All dose-response models will be re-fitted to the data (irrespective of statistical significance in the previous step) without any parameter assumptions to generate a set of new estimates of the model parameters. For all significant models, the generalised AIC will be calculated and reported in the displays. The final model will again be chosen as the model with the best fit (smallest AIC).

Estimates for each dose group will be calculated and will be based on the final dose-response model. No target dose criterion was pre-specified for the MRI-PDFF responder secondary endpoint.

Similar tables and figures will be used to display the results of the MCPMod analysis as were described in Section 7.4.1.

The above analysis will be performed on the TS using both actual maintenance treatment and planned maintenance treatment. In both cases, all post-baseline MRI-PDFF data within the Week 48 visit window will be used, both on- and off-treatment.



MRI-PDFF continuous secondary endpoints - main analysis

Descriptive summaries of liver fat content (measured in units of % using MRI-PDFF), as well as absolute and percentage changes from baseline, will be presented over time. These will be produced on the TS for both actual maintenance treatment and planned maintenance treatment, and will include both on- and off-treatment data in the relevant visit windows. The

descriptive summaries will be produced overall, and further split by presence of any type of diabetes (Yes, No).

An MMRM will be fitted to the absolute change from baseline in liver fat content values (measured in units of % using MRI-PDFF) at each scheduled week (calculated as per Section 6.7). The MMRM will include fixed effects for baseline liver fat content (%) as a continuous linear covariate, and treatment, presence of diabetes of any type [yes, no], baseline fibrosis score [F1, F2, F3], visit, treatment by visit interaction and baseline by visit interaction as factors. The Kenward-Roger approximation will be used to estimate the denominator degrees of freedom.

Unstructured covariance will be assumed to model the relationship between pairs of measurements taken at different weeks on the same patient. If the model fails to converge, all reasonable attempts will first be made to encourage convergence on the specified model with all scheduled weeks included. These attempts will include (but may not be limited to) first adjusting singularity options, then increasing the maximum number of convergence iterations, using the Fisher scoring algorithm for the early iterations, removing multi-threading options, providing starting values for the covariance parameters, and excluding the fixed effect for baseline fibrosis score from the model. Only if none of these options is successful will simpler covariance structures be considered, in which case these will be evaluated in a nested sequence, where each covariance matrix is a special case of all which have been tried before.

Adjusted means will be calculated from the MMRM using the observed margins approach, in which the contribution of model factors to the estimate is weighted proportionally to the presence of these factors in the data.

A summary table will be produced of the MMRM results at each time point, including adjusted means and their standard errors for each treatment group, and including adjusted mean differences from placebo for each Survodutide dose group and their 95% confidence intervals.

Adjusted means from the MMRM will also be displayed graphically over time to assess the onset of effect over time.

The above analysis will be performed on the TS for both actual maintenance treatment and planned maintenance treatment, and will include both on- and off-treatment data in the relevant visit windows.

The MMRM analysis specified above will be repeated for percentage change from baseline in liver fat content values (measured in units of % using MRI-PDFF).

MCPMod analysis will not be performed for the continuous MRI-PDFF secondary endpoints.

[redacted]

[redacted]

Liver biopsy responder secondary endpoint

Frequency tables will be produced of the liver biopsy responder secondary endpoint (one stage decrease in fibrosis stage from baseline after 48 weeks of treatment) in both of the following ways:

- TS, actual maintenance treatment, using all post-baseline liver biopsy data within the Week 48 visit window, both on- and off-treatment

- TS, planned maintenance treatment, using all post-baseline liver biopsy data within the Week 48 visit window, both on- and off-treatment.

The above frequency tables will be displayed similarly to those for the primary endpoint.

A logistic regression model will be fitted to the liver biopsy responder secondary endpoint without an intercept, and will include presence of diabetes of any type [yes, no], baseline fibrosis score [F1, F2, F3] and the dose group as factors. In the case that 0 events are observed in any combination of dose group and strata, the Firth bias reduction method will again be used.

A summary table will be produced of the logistic regression results, including estimated odds ratios for the comparison of each Survodutide dose against placebo and their 95% confidence intervals and associated p-values.

The above analysis will be performed on the TS for both actual maintenance treatment and planned maintenance treatment, and will include both on- and off-treatment data in the Week 48 visit window.

MCPMod analysis will not be performed for the liver biopsy responder secondary endpoint.

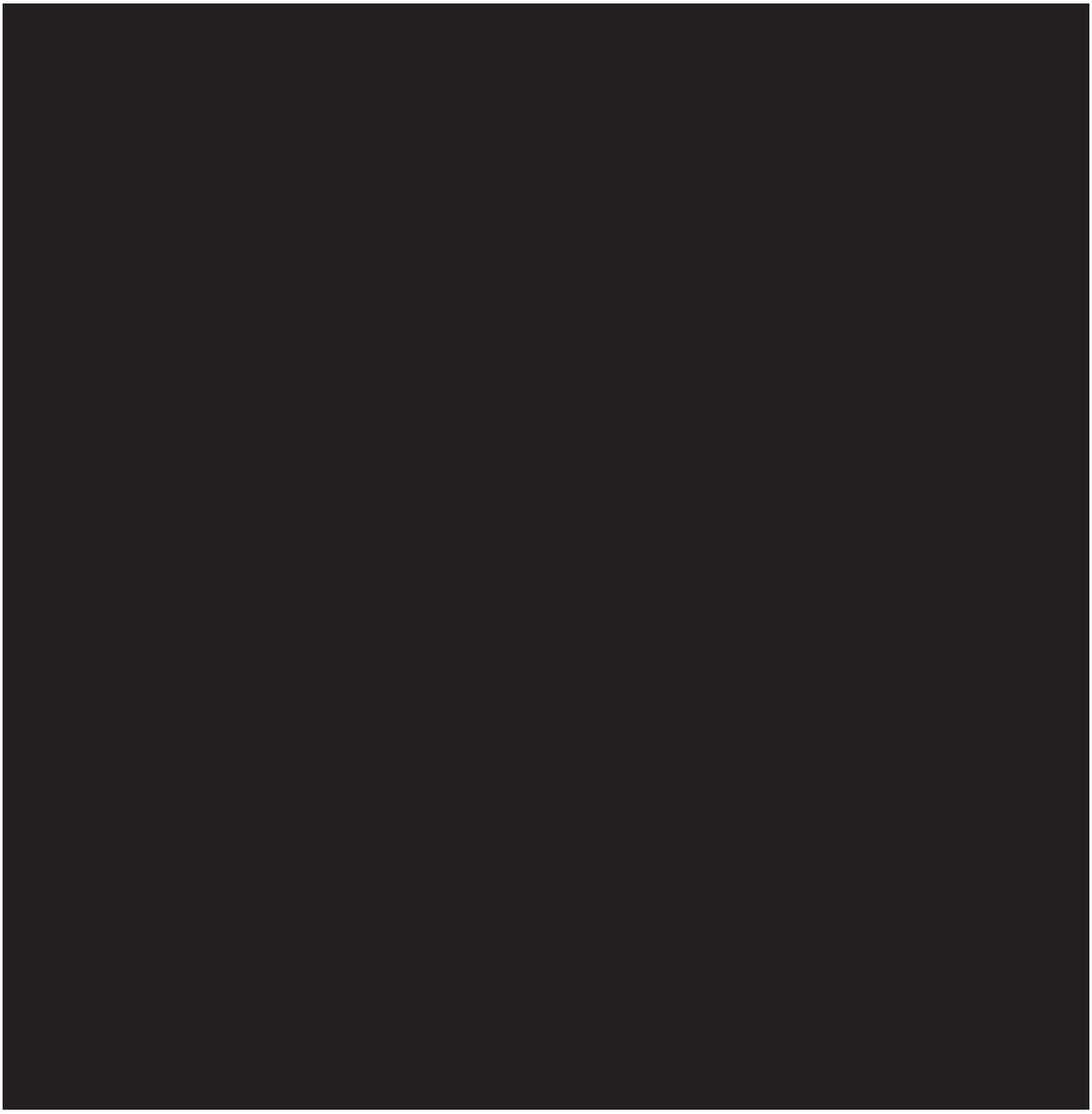
[redacted]

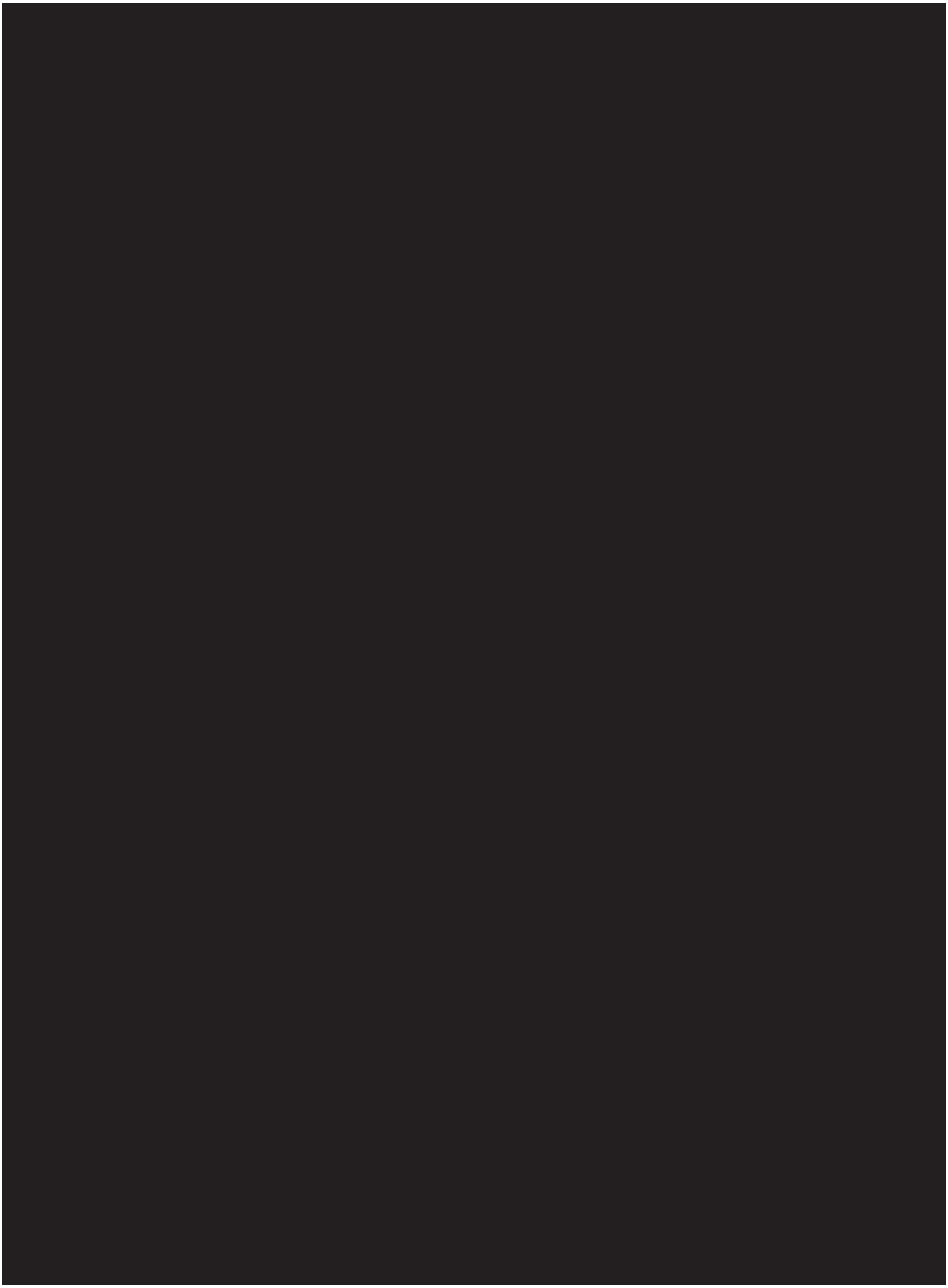
Liver biopsy continuous secondary endpoint

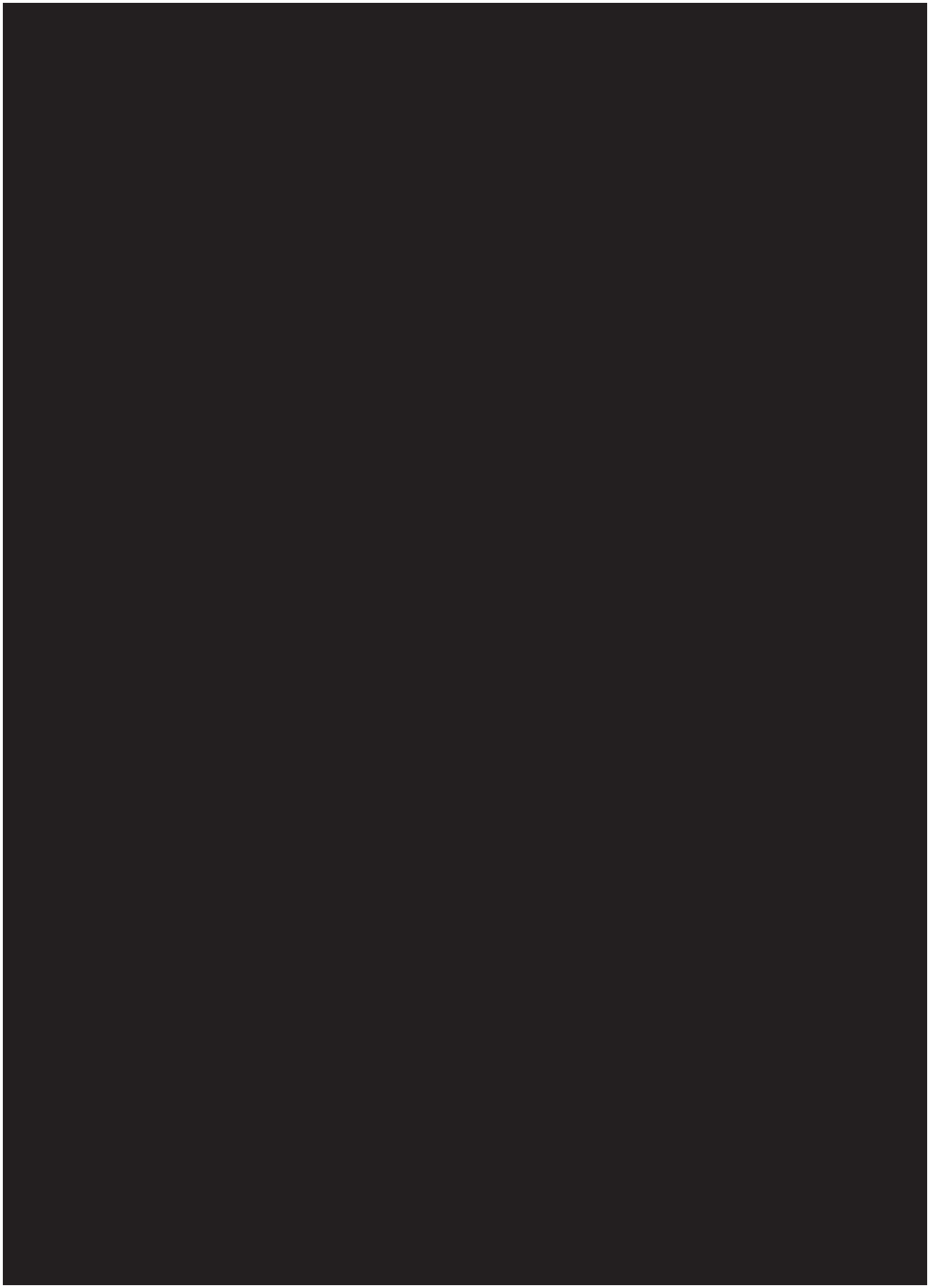
Descriptive summaries of NAS total score from the liver biopsy, as well as absolute change from baseline, will be presented at Week 48.

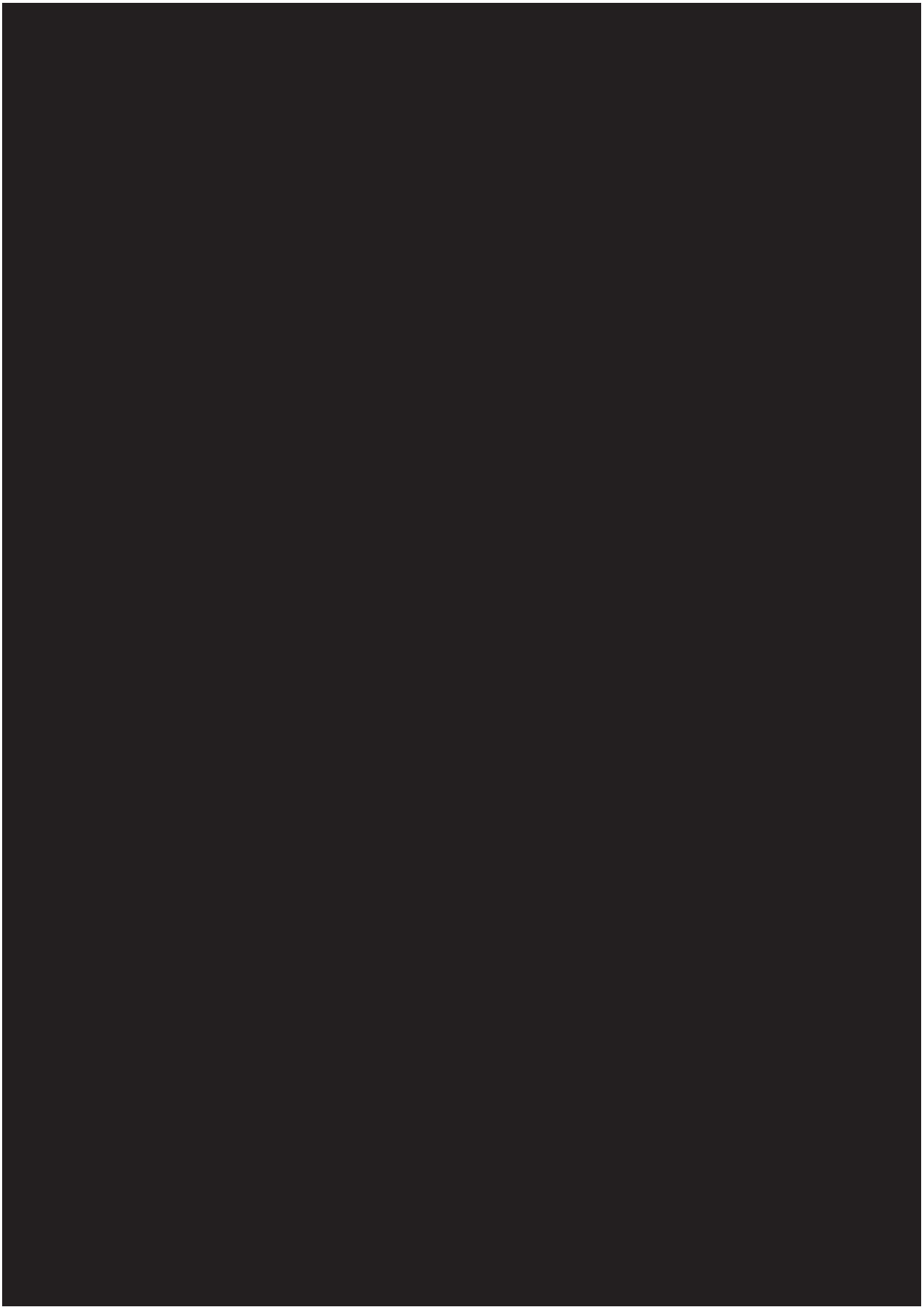
Frequency tables will also be produced for each category of the NAS total score, and for each of the 3 NAS sub-scores (steatosis, lobular inflammation and ballooning) at baseline and at Week 48.

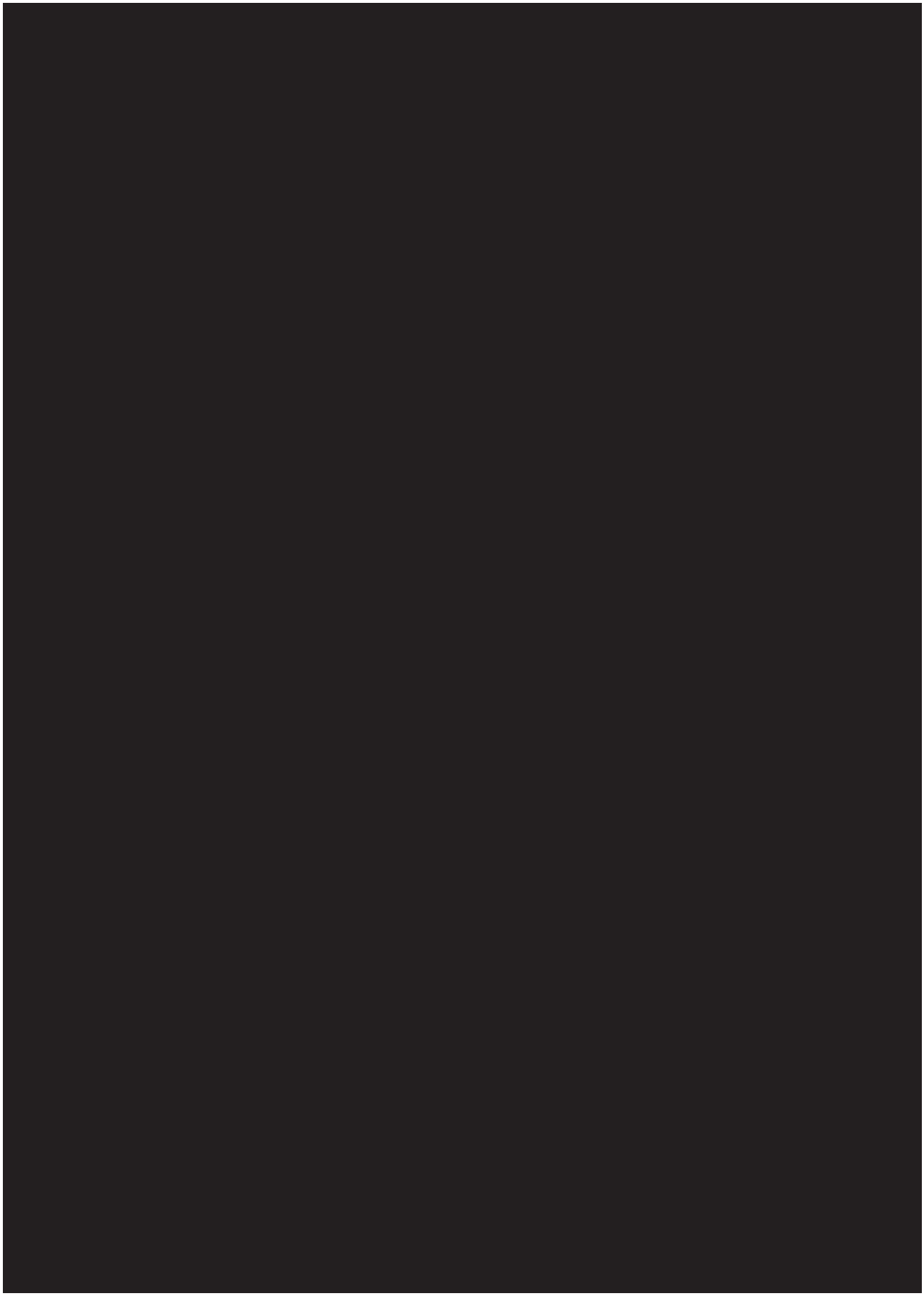
The descriptive summaries and frequency tables will be produced on the TS for both actual maintenance treatment and planned maintenance treatment, and will include both on- and off-treatment data in the relevant visit windows.

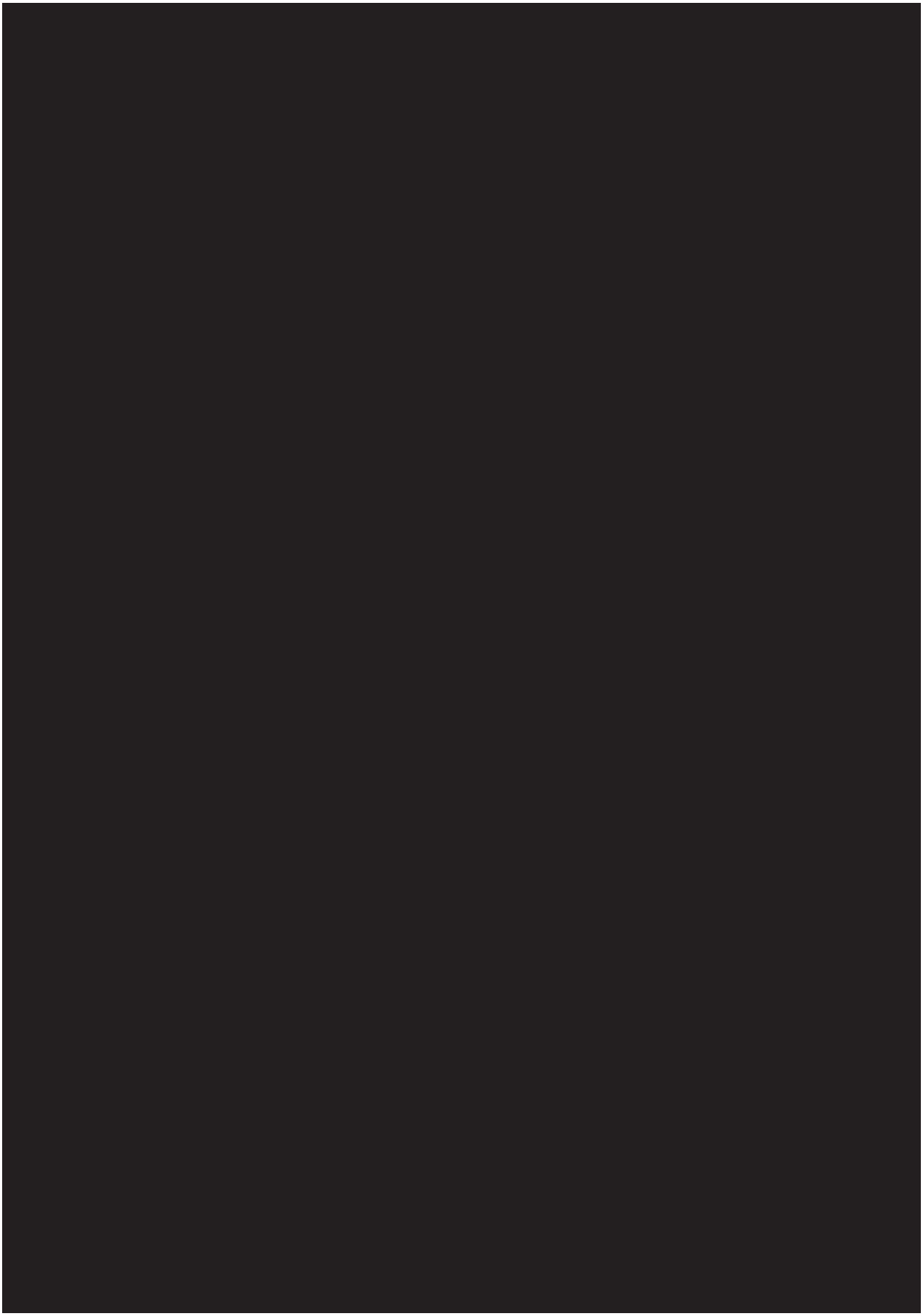


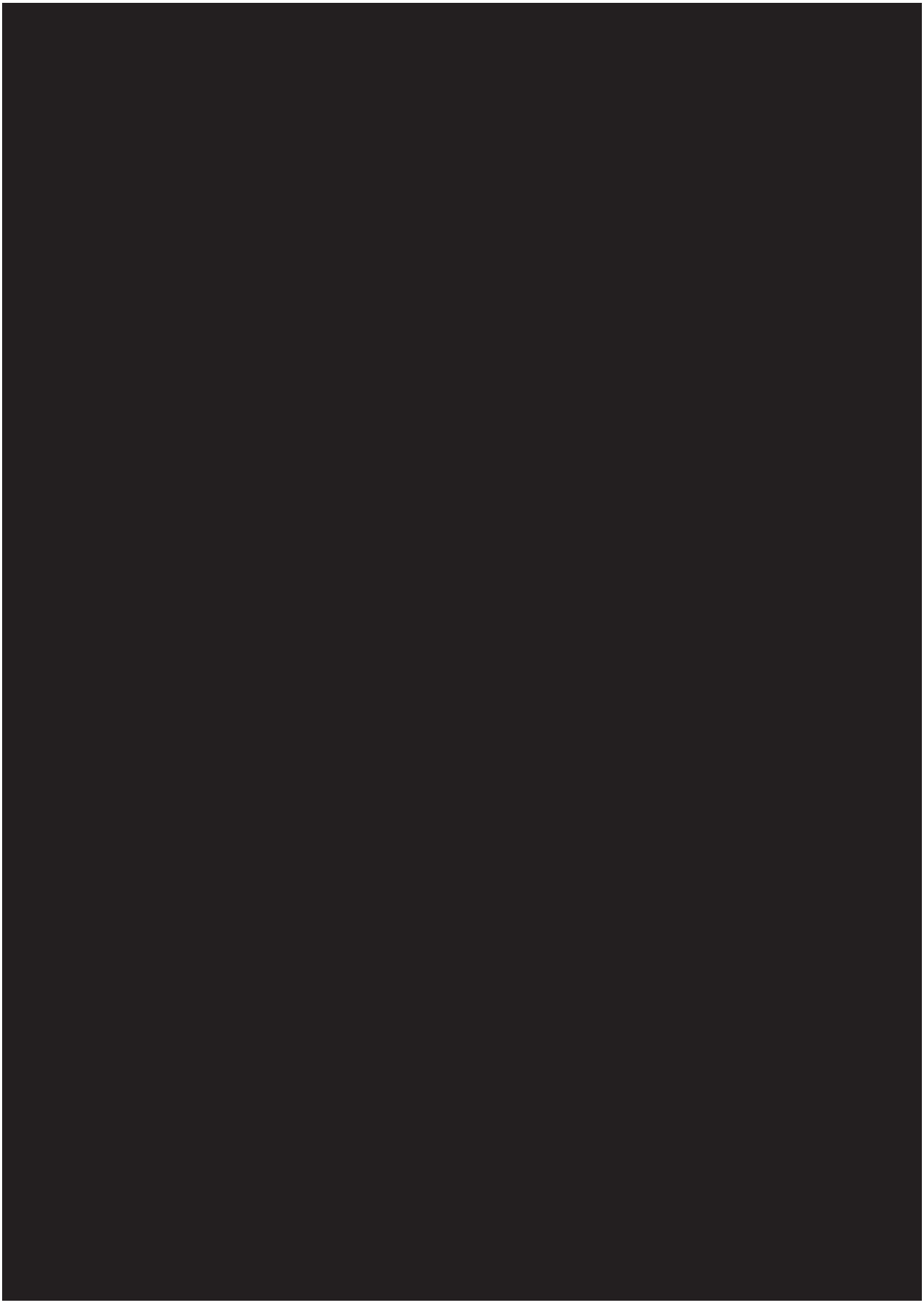












7.7 EXTENT OF EXPOSURE

Descriptive summary statistics will be produced of the extent of exposure, as defined in Section 5.4.3. A frequency table will also be presented of the extent of exposure by categories (both weekly and 4-weekly categorisations). The total patient-years exposure will also be summarised; the 28-day REP after the last dose of study treatment (see Section 6.1.6) will not be included for this purpose. These summaries will be produced overall on the TS using both actual maintenance treatment and planned maintenance treatment.

A frequency table will be produced of duration on study (regardless of whether the patient remained on-treatment) using the same 4-weekly categories as extent of exposure.

A descriptive summary and frequency table will be produced of the extent of exposure for patients who discontinued treatment early, using planned maintenance treatment only. This table will be produced in the following 4 ways (using weekly categorisations):

- All patients who prematurely discontinued treatment
- Patients who prematurely discontinued treatment due to AE
- Patients who prematurely discontinued treatment not due to AE
- Patients who prematurely discontinued treatment due to AE and at least one AE with SOC ‘Gastrointestinal disorders’ leading to treatment discontinuation.

7.8 SAFETY ANALYSIS

All safety analyses will be performed on the TS, including only on-treatment data (unless otherwise specified in Section 7.8.1). No hypothesis testing is planned for safety.

All specified summaries and analysis will be presented for both actual maintenance treatment and planned maintenance treatment.

7.8.1 Adverse Events

The analyses of adverse events will be descriptive in nature. All analyses of AEs will be based on the number of patients with AEs and not on the number of AEs.

The analysis of adverse events will be based on the concept of treatment emergent adverse events. AEs will be considered "on-treatment" if the AE start date meets the 28-day REP criterion (see Section 6.1.6). AEs occurring prior to the start of treatment will be assigned to "screening". AEs occurring more than 28 days after the last dose of study treatment will be assigned to "follow-up". Unless otherwise specified, on-treatment AEs will be pooled across dose escalation and maintenance periods.

An overall summary of adverse events will be presented for all patients in the TS. This summary will show the number and percent of patients with any AE, any investigator defined drug-related AE, any AE leading to the discontinuation of study medication, any AESI (as determined by the investigator), any serious AE (including reason for serious (death, life-threatening, disability/permanent damage, required or prolonged hospitalization, congenital anomaly/birth defect and other medically important events)), and any drug-related serious AE. The overall summary will be presented with the treatment period pooled across, and split into, dose escalation and maintenance periods, respectively.

AEs will be coded using the version of the MedDRA coding dictionary in force at final DBL. The frequency of patients with all AEs will be summarised by treatment, primary system organ class (SOC) and preferred term (PT). A patient with multiple on-treatment occurrences of the same preferred term meeting the criteria for inclusion in the table will be counted only once in these tabulations. SOCs will be sorted by descending overall frequency; PTs will be sorted by descending overall frequency (within SOC). The summary of all AEs will be presented with the treatment period pooled across, and split into, dose escalation and maintenance periods, respectively.

The AE summary by primary SOC and PT will also be repeated for each of the following, with the treatment period pooled across dose escalation and maintenance periods, only:

- AEs leading to treatment discontinuation
- Serious AEs
- Investigator-defined drug-related AEs
- Drug-related serious AEs.

An additional table will present the frequency of patients with AEs by worst CTCAE grade (1 to 5) and by primary SOC and PT. When no events meet the criteria for inclusion, the table will be shown as indicating no events of that type have occurred.

The following tables will be produced separately by baseline body weight subgroup (< 100 / ≥ 100 kg) and by presence of diabetes (yes/no):

- Overall AE summary
- Summary of all AEs by primary SOC and PT
- Summary of AEs leading to treatment discontinuation by primary SOC and PT.

For disclosure of AE data on ClinicalTrials.gov, the frequency of patients with non-serious AEs occurring with an incidence of greater than 5% (PT occurring in more than 5% of treated patients in one or more treatment groups) will be summarised by treatment, primary SOC and PT. The frequency of patients with serious AEs and drug related serious AEs will also be summarised respectively.

For disclosure of AE data in the EudraCT register, the frequency of patients with AEs, the frequency of patients with non-serious AEs with an incidence of greater than 5% (as above) and the frequency of patients with serious AEs will be summarised. Furthermore, the total number of treated patients by country and by age group will be summarised.

An exposure-adjusted summary of all AEs by primary SOC and PT will also be produced. For this purpose, the time at risk will be defined for specific AEs as follows:

- For patients with no events, the time at risk is the number of days on-treatment (inclusive), as defined in Section 6.1.6

- For patients with one or more event, the time at risk is the number of days from the start of treatment to the start of the first on-treatment event (inclusive).

The total time at risk (years) for a treatment group is the sum of the individual patient times at risk for that treatment group (days) divided by 365.25. Multiple occurrences of the same event for a particular patient will not be counted as separate events in this summary; a patient will either be considered to have no events of the type being summarised, or to have one or more occurrences of that event.

A frequency table will be produced of symptoms of injection site reaction (swelling, induration, heat, redness, pain, other) for the subset of patients with the event.

Additional trial-specific information captured in the eCRF for Grade 3 nausea will be listed only and not summarised.

Adverse Events of Special Interest (AESIs)

Hepatic injury is the only AESI pre-specified in the CTP. No further AESIs have been identified through increasing knowledge of the drug safety profile during the conduct of the trial.

AESIs as recorded by the investigator on the eCRF page will be summarised as part of the overall AE summary, described above.

Investigator-recorded AESIs will also be summarised by treatment, primary SOC and PT.

An additional table will present the frequency of patients with AEs possibly indicating hepatic injury, as defined by the four MedDRA subSMQ definitions given below:

- Liver related investigations, signs and symptoms (narrow subSMQ)
- Cholestasis and jaundice of hepatic origin (narrow subSMQ)
- Hepatitis, non-infectious (narrow subSMQ)

- Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (narrow subSMQ).

This table will be summarised by treatment, MedDRA subSMQ, primary system organ class and preferred term. Any preferred term which is classified under more than one subSMQ will be included under each relevant subSMQ.

Adjudicated Adverse Events

The number and percent of patients with each of the following confirmed adjudicated events (as defined in the CTP) will be presented:

- All-cause death
- Cardiovascular event
- Cerebrovascular disease
- Heart failure requiring hospitalisation
- Pancreatitis
- Neoplasm
- Thyroid mass requiring surgery.

It should be noted that "Heart failure requiring hospitalisation" is assessed by the same committee who are adjudicating cardiovascular events, and is recorded on the same form as all of the other cardiovascular event sub-categories (which were not distinguished in the same way in the CTP). Therefore the summary table will reflect the data actually captured.

7.8.2 Laboratory data

The analyses of laboratory data will be descriptive in nature and will be based on BI standards. Central laboratory data will be used for all displays described below, unless otherwise specified. Summaries of eGFR and UACR will be included as part of the routine summaries of laboratory variables.

For continuous safety laboratory variables where reference ranges exist, normalised values will be derived (i.e. transformation to both a standard unit and to a standard reference range). Descriptive statistics will be presented over time for absolute values and changes from baseline in each of the laboratory variables where normalised values are possible. All scheduled weeks will be presented as well as last value, minimum value and maximum value on-treatment.

For continuous safety laboratory variables where reference ranges exist, a shift table will be produced summarising the shifts from baseline to the last value on treatment, and displaying categories "Low", "Normal" and "High" with respect to the relevant reference range. Standardised values (i.e. transformation to a standard unit only) will be used for this purpose.

For categorical safety laboratory variables, a shift table will be produced summarising the shifts from baseline to the last value on treatment, and displaying each relevant category. For any semi-quantitative variables (i.e. those recorded as either "Normal" or with a continuous numerical value otherwise), the shift table will display categories "Normal" and "Abnormal"

only, without further categorisation of the continuous values into different degrees of abnormality.

Potentially clinically significant abnormalities will be identified using BI standard rules. A listing containing these rules (including for each variable whether a low range applies, or a high range applies, or both) will be included. A frequency table will be used to summarise the number and percentage of patients with potentially clinically significant abnormalities, for each variable where a BI standard rule exists. Patients with an abnormal value at baseline will be presented separately on this table.

A separate listing will present potentially clinically significant abnormal lab values; for each functional lab group all patient's laboratory values will be listed, if there exists at least one value with clinically significant abnormality within the group.

The following additional investigations will be made for elevated liver enzymes:

A frequency table will be produced of the following categories. A patient will be included in each relevant category, defined using the maximum post-baseline ALT and AST values observed on-treatment (Section 6.1.6), regardless of the baseline AST and ALT values:

- ALT or AST ≥ 3 x ULN
- ALT or AST ≥ 5 x ULN
- ALT or AST ≥ 8 x ULN
- ALT or AST ≥ 10 x ULN
- ALT or AST ≥ 20 x ULN
- ALT or AST ≥ 3 x ULN with total bilirubin ≥ 2 x ULN
- ALT or AST ≥ 3 x ULN with INR ≥ 1.5
- ALT or AST ≥ 3 x ULN with either total bilirubin ≥ 2 x ULN or INR ≥ 1.5

In the above categories, each occurrence within a patient of a post-baseline elevation ALT or AST ≥ 3 x ULN will be used as a trigger for the determination of the elevation of total elevation of bilirubin ≥ 2 x ULN or INR ≥ 1.5, as applicable. The proposed time interval for the follow-up of total bilirubin or INR is 30 days after the occurrence of the post-baseline ALT or AST elevation.

The above frequency table will be repeated using the same categories, but restricted further to patients with a normal baseline value for ALT and AST, defined as both ALT < ULN and AST < ULN.

An additional frequency table will be produced for patients who have an abnormal baseline value for ALT or AST, defined as ALT ≥ ULN or AST ≥ ULN (or both), and who fall into each of the following categories:

- Both baseline ALT and AST < 2 x ULN and post-baseline ALT or AST > 5 x baseline value
- At least one of baseline ALT or AST ≥ 2 x ULN and < 5 x ULN and post-baseline ALT or AST > 3 x baseline value

- At least one of baseline ALT or AST ≥ 5 x ULN and post-baseline ALT or AST > 2 x baseline value
- Post-baseline ALT or AST > 2 x baseline value with total bilirubin ≥ 2 x ULN or increase in INR > 0.2.

7.8.3 Vital signs

Descriptive statistics will be presented over time for absolute values and changes from baseline in each of the vital signs variables, and will include display of pre- and post-dose at each scheduled week.

7.8.4 ECG

Abnormal findings will be reported as baseline conditions (at screening) or as AEs (during the trial) if judged clinically relevant by the investigator.

The ECG recordings will be centrally evaluated and rated as normal, abnormal, or not evaluable.

All derivations relating to 12-lead ECG variables are given in Section 5.4.1.

Occurrences of values above thresholds will be flagged in the listings. For QTcB and RR, only listings will be provided.

Quantitative variables

Descriptive statistics will be presented over time for absolute values and absolute changes from baseline in the following quantitative ECG variables: QTcF, HR, QT, QRS, PR.

Graphical summaries over time will also be included.

Categorical variables

Frequency tables will be provided for the categorical variables described below, which are determined from the specified quantitative ECG variables:

- New onset (meaning that this or a higher category was not present any time at baseline) of maximum QTcF interval > 450 to 480 msec, > 480 to 500 msec, or > 500 msec on treatment.
- New onset (not present any time at baseline) of uncorrected QT interval > 500 msec at any time on treatment
- Maximum change from baseline of QTcF ≤ 30 msec, > 30 to ≤ 60 msec, or > 60 msec on treatment.
- Maximum change from baseline of QT ≤ 60 msec, or > 60 msec on treatment.

For assignment of a particular patient to one of the above categories, all available time points on-treatment will be considered.

The occurrence of any of the following will be considered as "notable findings" and summarised using frequency tables:

- New onset (not present any time at baseline) of uncorrected QT interval > 500 msec at any time on treatment
- New onset of QTcF interval > 500 msec at any time on treatment
- Change from baseline of QTcF > 60 msec at any time on treatment
- Percentage change from baseline of HR ≥ 25%, when corresponding on-treatment value of HR is > 100 beats/min, or percentage change from baseline of HR ≤ - 25%, when corresponding on-treatment value of HR is < 50 beats/min, at any time on-treatment
- Percentage change from baseline of PR ≥ 25%, when corresponding on-treatment value of PR interval is > 200 msec, at any time on treatment
- Percentage change from baseline of QRS ≥ 10%, when corresponding on-treatment value of QRS duration is > 110 msec, at any time on treatment.

Frequency tables will also include morphological findings (determined and categorized based on SDTM terminology) that might be attributable to treatment. In particular, new onsets of findings not present at baseline will be explored. A morphological finding observed on treatment that was not reported at baseline will be categorized as a 'new onset' of this finding.

For all patients with any notable finding in ECG intervals, a separate listing will be created as end-of-text display, and the corresponding time profiles will be shown.

A shift table will be produced for QTcF, summarising the shifts from baseline to the maximum value on treatment, and displaying categories "≤ 450 msec", "> 450 to 480 msec", "> 480 to 500 msec and "> 500 msec".

A scatter plot of QTcF at baseline and the maximum change from baseline (based on all on-treatment values as defined in Section 6.1.6, and the maximum evaluated regardless of whether these were selected based on visit windows described in Section 6.7) will be produced. The plot will include diagonal reference lines for absolute QTcF equal to 450 msec, 480 msec and 500 msec, as well as horizontal reference lines for QTcF maximum changes from baseline equal to 30 msec and 60 msec.

Appropriateness of heart rate correction methods of QT interval

To evaluate the appropriateness of the heart rate correction methods, the slope of the relationship of the (untransformed) QTcF interval versus RR interval will be estimated for off-drug values and for each Survodutide maintenance dose group separately by applying a random coefficient (random slope and intercept) model. Off-drug values for this purpose are defined as only values (either baseline or on-treatment values as defined in Section 6.1.6) from patients randomised to placebo, since patients randomised to Survodutide dose groups are expected to have a single ECG at baseline only, and therefore cannot contribute a slope for estimation of an off-drug random effect for this. For the Survodutide maintenance dose group analyses, only on-treatment values will be used, and these will be included regardless of Survodutide dose (i.e. data from the dose escalation period will be included). The model will be fitted using the QTcF and RR pairs at each time point, using all values within the visit windows defined in Section 6.7. A scatter plot of QTcF versus RR will be produced, and will

include the overall regression line obtained from the model. The population estimate of the slope together with its two-sided 95% confidence interval will be displayed in a footnote to the plot.

The slope of the relationship of the log-transformed (uncorrected) QT interval versus log-transformed RR interval will be estimated for off-drug values and for each Survodutide maintenance dose group separately by applying a random coefficient (random slope and intercept) model similar to above, with off-drug values and data selection defined in the same way. A scatter plot of log(QT) versus log(RR) will be produced, and will include the overall regression line obtained from the model. The population estimate of the slope together with its two-sided 95% confidence interval will be displayed in a footnote to the plot.

7.8.5 Other

Frequency tables will be produced of the information in the C-SSRS questionnaire over time.

7.9 ANALYSIS OF COVID-19 IMPACT

To assess the impact on patients' safety and drug efficacy in this trial, the following analyses are planned:

The frequency of patients with missed relevant visits or early discontinuation from study treatment or study due to COVID-19 and related PDs will be listed and analysed descriptively.

Summary tables and listings of AEs of SARS-CoV-2 infection will be produced. The tables will include an overall summary, a summary by primary SOC and PT, a summary of relevant AEs which led to early treatment discontinuation, and a summary of relevant serious AEs. For this purpose, "SARS-CoV-2 infection" is defined by the broad BIcMQ of SARS-CoV-2 infections with search ID 32010051. The broad BIcMQ differs from the narrow BIcMQ only by the inclusion of the PT for "Suspected COVID-19".

AEs related to SARS-CoV-2 infection will also be summarised and listed. For this purpose, "related to SARS-CoV-2 infection" is defined by the MedDRA broad SMQ for COVID-19.

8. TIMEPOINT OF RELEASE OF TREATMENT INFORMATION

The treatment information will be released to unblind the trial database after the last patient has completed their End-of-Study/Follow-up visit and all data has been entered and cleaned as defined in the “Data Ready to be Unblinded and/or Final Trial Closure Notification” (RUN) form.

9. REFERENCES

1.	*CPMP/ICH/363/96:* "Statistical Principles for Clinical Trials", ICH Guideline Topic E9, Note For Guidance on Statistical Principles for Clinical Trials, current version.
2.	*BI-KMED-BDS-MAN-0011:* "Link to Statistical Position Papers Homepage", current version; KMed
3.	Firth D. Bias reduction of maximum likelihood estimates. Biometrika 80, 27–38 (1993) [R20-2291]
4.	Bornkamp B, Pinheiro J, Bretz F Package 'DoseFinding' (February 19, 2015). http://cran.r-project.org/web/packages/DoseFinding/DoseFinding.pdf (access date: 28 April 2015); Comprehensive R Archive Network (2015). [R15-2001]
5.	Pinheiro JC, Bornkamp, B, Bretz F: Design and analysis of dose-finding studies combining multiple comparisons and modeling procedures. Journal of Biopharmaceutical Statistics 2006; 16: 639-656.[R10-1424]
6.	Pinheiro JC, Bornkamp, B, Glimm, E, and Bretz, F: Model-based dose-finding under model uncertainty using general parametric models. Statistics in Medicine 2014; 33(10); 1646–1661. [R15-4293]

11. HISTORY TABLE

Table 11: 1 History table

Version	Date (DD-MMM-YY)	Author	Sections changed	Brief description of change
1.0	11DEC2023	[redacted]	None	This is the final TSAP
2.0	19JAN2024		Section 5.1 Section 6.7	This is the revised TSAP. The baseline definition for all biopsy endpoints, including the primary endpoint, has been changed for patients in China, to use the digital reading (1st screening reading) instead of the physical reading (2nd screening reading). This has been done in response to potential analysis model and endpoint derivation issues which were identified from review of these data (transferred to the database 18DEC2023), specifically a thorough consideration of all the implications of physical readings (2nd screening readings) which yielded values outside the eligibility criteria ranges.