Statistical Analysis Plan	
Official Title:	A Randomized, Double-Blind Study to Evaluate the Efficacy and Safety of Tideglusib Versus Placebo for the Treatment of Children and Adolescents with Congenital Myotonic Dystrophy (REACH CDM)
NCT Number:	NCT03692312
Document Date:	Main Statistical Analysis Plan – 22 February 2023 Main Statistical Analysis Plan Addendum – 22 August 2024

A Randomized, Double-Blind Study to Evaluate the Efficacy and Safety of Tideglusib Versus Placebo for the Treatment of Children and Adolescents with Congenital Myotonic Dystrophy (REACH CDM)

Main Statistical Analysis Plan

Version: Final 2.0

Date: 22nd February 2023

Author

[redacted]

[redacted]

22-Feb-2023

Approved By (AMO Pharma Ltd)

[redacted]

[redacted]

22-Feb-2023

[redacted]

22-Feb-2023

Modification History

Version	Change History	Reason	Date
1.0	First Version	N/A	01DEC2021
2.0	Second Version	Updated following FDA feedback on version 1.0 and update to sample size due to SSRE.	20FEB2023

Contents

LIST OF ABBREVIATIONS

ABI-C	Autism Behaviour Inventory-Clinician
AE	Adverse Event
ANCOVA	Analysis of Covariance
AR(1)	Auto-Regressive with a Lag of 1
ASD	Autism Spectrum Disorder
BDRM	Blinded Data Review Meeting
BMI	Body Mass Index
BP	Blood Pressure
CC-MDHI	Congenital and Childhood Myotonic Dystrophy Health Index
CC-CDM1-RS	Caregiver-Completed Congenital DM1 Rating Scale
CDM1-RS	Clinician-Completed Congenital DM1 Rating Scale
Congenital DM1	Congenital Myotonic Dystrophy
COVID-19	Coronavirus Disease 2019
CGI-I	Clinical Global Impression Clinical Global Impression – Improvement Scale
CGI-S	Clinical Global Impression- Severity Scale
CI	Confidence Interval
DCCS	Dimensional Change Card Sort Test
DOB	Date of Birth
DSMC	Data Safety Monitoring Committee
DXA	Dual-energy X-ray Absorptiometry
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
EOT	End of Treatment
FAS	Full Analysis Set
FCS	Fully Conditional Specification
FDA	Food and Drug Administration
FT3	Free Triiodothyronine
FT4	Free Thyroxine
[redacted]	[redacted]
HbA1c	Gycated Hemoglobin
ICH	International Conference on Harmonisation

ITT	Intent-to-treat
LLOQ	Lower Limit of Quantification
LS	Least Squares
MAR	Missing at Random
MCMC	Markov Chain Monte Carlo
MNAR	Missing not at Random
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation
MMRM	Mixed Model Repeated Measures
NIH	National Institutes of Health
NTEAE	Non-Treatment Emergent Adverse Event
PDC	Protocol Deviation Criteria
PK	Pharmacokinetic
PKAS	Pharmacokinetics Analysis Set
PPS	Per Protocol Set
PPVT	Peabody Picture Vocabulary Test
PSMT	Picture Sequence Memory Test
PT	Preferred Term
QTcB	QT Interval – Bazett's Correction
QTcF	QT Interval – Fridericia's Correction
REML	Restricted Maximum Likelihood
RIS	Run-In Set
RNA	Ribonucleic Acid
SAE	Serious Adverse Event
SAF	Safety Analysis Set
SAP	Statistical Analysis Plan
SD	Standard Deviation
SE	Standard Error
SOC	System Organ Class
TEAE	Treatment Emergent Adverse Event
TSH	Thyroid Stimulating Hormone
ULN	Upper Limit of Normal
VAS	Visual Analog Scale
WOCF	Worst Observation Carried Forward
WHO Drug	World Health Organization Drug Dictionary

1 INTRODUCTION

This document details the statistical analysis of the data that will be performed for the AMO Pharma Ltd study: AMO-02-MD-2-003.

The proposed analysis is based on the contents of the Final Version 8.0 of the protocol (dated 15-June-2022). Additionally, the following guidance documents have been consulted: ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials (dated May 2021), Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products (draft guidance, dated May 2021), Rare Diseases: Common Issues in Drug Development Guidance for Industry (dated February 2019), and Adaptive Design Clinical Trials for Drugs and Biologics Guidance for Industry (dated December 2019).

In the event of future amendments to the protocol, this statistical analysis plan (SAP) may be modified to account for changes relevant to the statistical analysis.

The table, listing and figure shells are supplied in a separate document.

2 STUDY OBJECTIVES AND DESIGN

AMO-02-MD-2-003 is a randomized double-blind study of weight adjusted 1000 mg tideglusib versus placebo across a [redacted]-week treatment period. The subjects are children between the ages of 6 and 16 years with Congenital Myotonic Dystrophy (Congenital DM1). Approximately 56 children will be randomized into the study.

2.1 Study Objectives

The primary objective of the study is to evaluate the efficacy, from baseline to end of treatment, of weight adjusted 1000 mg tideglusib compared to placebo in children and adolescents with Congenital DM1. The primary efficacy measure is the Clinician-Completed Congenital DM1 Rating Scale (CDM1-RS).

The secondary objectives of the study are to evaluate weight adjusted 1000mg tideglusib versus placebo in children and adolescents with Congenital DM1 for the following:

- Safety and tolerability.
- Efficacy, from baseline to end of treatment, as measured by clinician completed rating scales, caregiver-completed rating scales, functional assessments, and biomarker/physiological assessments.
- Blood pharmacokinetics of tideglusib and its main metabolite (NP04113) after repeat dosing.
- Consistency of telehealth data and in-clinic data for the CDM1-RS and Clinical Global Impression (CGI) rating scales.

2.2 Study Estimands

Following the adoption by FDA of ICH E9 R1 addendum on estimands and sensitivity analyses, the following estimands have been defined.

The estimand for the primary efficacy endpoint is the difference between group means in the change from baseline in total score on the Clinician-Completed Congenital DM1 Rating Scale (CDM1-RS) at end of treatment (EOT).

Estimand components (Combined approach):

A. The population is defined through appropriate inclusion/exclusion criteria to reflect the targeted patient population.
B. The variable is the change from baseline in CDM1-RS total score at EOT.
C. The intervention effect of subjects who discontinue treatment or discontinue the study due to AE/ death will be calculated based on available data up to the time of study discontinuation.
D. The population-level summary measure is the difference between adjusted group means in the change from baseline in the CDM1-RS at EOT.

The introduction of new physical therapy, use of permitted concomitant medication and non-compliance to protocol (e.g., poor treatment adherence, use of prohibited medication and other protocol deviations) are assumed to be well-balanced events across both arms and should not affect the estimation of the treatment effects. Therefore, in the first instance all observed data will be used regardless of the occurrence of these intercurrent events. In-clinic visits are of primary interest. If a subject's in-clinic visit is unavailable due to COVID-19, the data will be imputed with the corresponding telehealth assessment.

Missing data due to early study discontinuations or missed visits (where both in-clinic and telehealth are missed) are assumed to be Missing-At-Random (MAR) and will not be imputed. Instead, a direct likelihood approach such as a Mixed Model Repeated Measures (MMRM) will be used to analyze the available data.

Sensitivity and supplementary analyses will assess the robustness of the primary analysis to these chosen strategies and will address the occurrence and potential imbalance of intercurrent events, deaths and COVID-19 impacts between treatment groups.

The secondary estimand for the study is CGI-I at EOT:

The estimand for the key secondary efficacy endpoint is the difference between treatment means in the observed CGI-I score at EOT.

Estimand components (Combined approach):

A. The population is defined through appropriate inclusion/exclusion criteria to reflect the targeted patient population.
B. The variable is the observed values of the CGI-I score at EOT.
C. The intervention effect of subjects who discontinue treatment or discontinue the study due to AE/ death will be calculated based on available data up to the time of study discontinuation.
D. The population-level summary measure is the difference in adjusted means between treatment groups at EOT.

The introduction of new physical therapy, use of permitted concomitant medication and non-compliance to protocol (e.g., poor treatment adherence, use of prohibited medication and other protocol deviations) are assumed to be well-balanced events across both arms and should not affect the estimation of the treatment effects.

Therefore, in the first instance all observed data will be used regardless of the occurrence of these intercurrent events. In-clinic visits are of primary interest. If a subject's in-clinic visit is unavailable due to COVID-19, the data will be imputed with the corresponding telehealth assessment.

Missing data due to early study discontinuations or missed visits (where both in-clinic and telehealth are missed) are assumed to be Missing-At-Random (MAR) and will not be imputed. Instead, a direct likelihood approach such as a Mixed Model Repeated Measures (MMRM) will be used to analyze the available data.

Sensitivity and supplementary analyses will assess the robustness of the primary analysis to these chosen strategies and will address the occurrence and potential imbalance of intercurrent events, deaths and COVID-19 impacts between treatment groups.

2.3 Study Endpoints

The study endpoints (relating to the estimand variables in section 2.2 above) are as follows. Note that the terms 'endpoint' and 'variable' may be used interchangeably throughout this SAP to refer to the measurement of interest:

The primary efficacy endpoint is:

- Change from baseline in Clinician-Completed Congenital DM1 Rating Scale (CDM1-RS) to end of treatment

The key secondary efficacy endpoint is:

- Clinical Global Impression - Improvement Scale (CGI-I) (also referred to as the CGI-C*)* at end of treatment

The secondary endpoints are:

- Change from baseline in Top 3 Caregiver Concerns VAS score to end of treatment
- Change from baseline in Caregiver Completed Congenital DM1 Rating Scale (CC-CDM1-RS) to end of treatment
- Change from baseline in Clinical Global Impression - Severity Scale (CGI-S) to end of treatment
- Change from baseline in Clinician-Completed Congenital DM1 Rating Scale (CDM1-RS) - Independent Central Rater Score to end of treatment
- Clinical Global Impression - Improvement Scale (CGI-I) - Independent Central Rater Score at end of treatment
- Change from baseline in Clinical Global Impression – Severity Scale (CGI-S) – Independent Central Rater Score to end of treatment
- Change from baseline in 10-meter walk-run test (preferred speed and fastest speed) to end of treatment

The exploratory endpoints are:

- Change from baseline in DXA Scan measurement of total body lean/muscle mass to end of treatment

- Change from baseline in measurement of lip strength (via lip force meter) to end of treatment
- Change from baseline in Congenital and Childhood Myotonic Dystrophy Health Index (CC-MDHI) Parent Proxy Instrument to end of treatment
- Change from baseline in Autism Behavior Inventory - Clinician (ABI-C) to end of treatment
- Change from baseline in [redacted], and [redacted] standard scores of the Vineland Adaptive Behavior Scale – [redacted] to end of treatment
- Change from baseline in quantitative myometric measure of hand grip strength to end of treatment
- Change from baseline in NIH Toolbox Cognition Battery: Dimensional Change Card Sort Test to end of treatment
- Change from baseline in NIH Toolbox Cognition Battery: Picture Sequence Memory Test to end of treatment
- Change from baseline in Peabody Picture Vocabulary Test (PPVT) to end of treatment
- [redacted] evels
- [redacted] sample
- [redacted]
- Serial blood pharmacokinetics of tideglusib

The Safety endpoints are:

- The incidence of Adverse events (AEs), including serious adverse events (SAEs), between Screening and end of treatment. The incidence will also be assessed during a [redacted] follow-up period after discontinuation of the study drug for those subjects not participating in the extension study AMO-02-MD-2-004.
- The incidence of abnormal findings in objective assessments (e.g. laboratory values, ECGs, vital signs and bone mineral density) between screening and end of treatment. The incidence of abnormal findings in objective assessments will also be assessed during a [redacted] follow-up period after discontinuation of the study drug for those subjects not participating in the extension study AMO-02-MD-2-004.

2.4 Study Design

This is a phase 2/3, randomized, double-blind, placebo-controlled, parallel group, superiority study of weight adjusted dose 1000 mg/day tideglusib versus placebo in subjects aged 6 to 16 years with Congenital Myotonic Dystrophy (Congenital DM1) across a [redacted] treatment period.

Approximately 56 children will be randomized 1:1 into the study in a double-blind manner to either tideglusib or placebo such that each treatment group will have approximately 28 children (assuming a 10-13% drop-out rate) contributing data to the

primary efficacy analysis. The randomisation scheme will be stratified by age at time of screening [redacted]). Subjects are planned to be recruited from approximately 11 sites in the US, UK, Canada, Australia and New Zealand.

The study will have 5 distinct phases:

- Screening (Weeks [redacted] where subjects are screened to ensure adherence to eligibility criteria.
- [redacted] run-in [redacted]) for all subjects. Subjects and their caregiver are dispensed [redacted] and provided with instructions on when and how to administer the medication. Subject diaries will be provided to record the date and time of dose, any issue with taking the dose, information on the subject's food intake before and after dosing. Diary entries will be reviewed by the investigator site at each visit.
- Double-blind dose [redacted] (Weeks [redacted] At visit [redacted] (Baseline) subjects are randomized 1:1 to the weight-adjusted 1000 mg tideglusib or a matched placebo. Subjects randomized to placebo remain on that treatment. Subjects randomized to weight adjusted 1000 mg tideglusib will take tideglusib at a weight adjusted 400 mg dose level from [redacted] then they will take tideglusib at a weight adjusted 600 mg dose level from weeks [redacted] and thereafter they will take tideglusib at the weight adjusted 1000 mg dose level from weeks [redacted]
- Double-blind maintenance (Weeks [redacted]): Subjects receive a fixed dose of 1000 mg weight-adjusted tideglusib or matched placebo for [redacted] weeks from weeks [redacted] The final study assessment for subjects who proceed into the extension study AMO-02-MD-2-004 is Week [redacted]
- Follow-up Period (Weeks [redacted]): Will be for a [redacted] period after end of treatment for subjects not participating in the extension study AMO-02-MD-2-004. Adverse Events will be elicited at [redacted] and [redacted] ([redacted]) post end of treatment visit via telephone by a member of the study team.

The overall expected duration of subject participation in AMO-02-MD-2-003 is [redacted] weeks.

[redacted]

2.5 Visit Structure

The visit structure and scheduled assessments are detailed in Table 1 of the protocol.

3 SAMPLE SIZE

Twenty-five subjects per group will generate 80% power to detect the effect size of 0.82 at the 0.05 two-sided significance level. This sample size will also be sufficient to generate 70% power to detect the effect size of 0.6 at the 0.1 two-sided significance level. Accounting for a 10%-13% drop-out rate, approximately 28 subjects per group will be enrolled.

No data are available on the CDM1-RS instrument. Based on the preliminary data observed in the AMO-02-MD-2-001 study, the assumed effect sizes look reasonable. The effect size assumptions are partially based on the Clinician VAS total score changes from baseline to end of study (6 points on drug vs. 1 point on placebo with the common standard deviation of 5.8). Based on these and other data, [redacted] and [redacted] subjects per group will be sufficient to generate 90% and 80% power at the 0.05 two-sided significance level. Based on these calculations, the suggested sample size of 56 subjects total is adequate to detect the treatment effect or at least a consistent trend.

A blinded sample size re-estimation will be performed, per FDA guidance on adaptive designs (Section IV, Adaptive designs based on non-comparative data), once [redacted] of subjects have been enrolled and have completed Visit [redacted] All study subjects, investigators, site staff and Sponsor's staff and delegates (including the statistical consultant and SQN staff) will remain blinded to treatment assignments.

The blinded sample size re-estimation will be conducted based on the observed standard deviation in the change from baseline in the clinician-completed CDM1-RS at EOT in the ITT population in those subjects who have completed Visit [redacted]. Summary statistics (mean, median, standard deviation, minimum and maximum) and histograms of the observed values and change from baseline will be provided by visit up to Visit [redacted] with all subjects being presented under an 'Overall' treatment group. This summary will provide an estimate of the pooled standard deviation for evaluation. Similar summaries will be provided for the clinician-completed CGI-I and the Top 3 Caregiver Concerns VAS score. The blinded external statistical consultant will then conduct the sample size scenarios.

No data on treatment effect and variability based on the CDM1-RS instrument are available; the estimates of mean changes from baseline are not available for placebo, or for the active treatment arm. Three initial scenarios are pre-specified here (treatment effect Δ=5.0, Δ =3.6, and Δ =2.1). However, the clinical relevance of these assumptions is unknown. Additional scenarios will be based on the observed standard deviation and additional assumed treatment differences. Scenarios involving the CGI-I and/or the Top 3 Caregiver Concerns VAS score may also be assessed. The decisions regarding corresponding sample size increases will be made based on these estimates and other considerations.

Dependent upon the blinded estimate of standard deviation, the sample size may be increased to ensure 80% power for a chosen treatment effect. Any increase in sample size will be limited to 2 times the maximum of the initial sample size. Additional (not pre-specified) scenarios may be requested by the Sponsor after pre-specified scenarios have been reviewed based on various assumptions regarding effect size.

Sample size update due to SSRE

Based on the blinded SSRE, the planned sample size is sufficient to detect the treatment difference between active and control arms based on the primary endpoint with at least 80% power at the 0.05 two-sided significance level. However, the variability of the key secondary endpoint was slightly greater than that originally assumed. Respectively, it was recommended, if operationally feasible, to increase the total sample size to 66 subjects. Further details are in Appendix 2.

As it was not operationally feasible to enroll 66 subjects, 53 subjects were randomized. Sample size calculations based on SSRE results supported this operational decision.

4 RANDOMIZATION

Randomization will occur in a 1:1 manner (e.g. 1000 mg weight adjusted tideglusib: placebo) such that each treatment group will have approximately 28 children and adolescents contributing data to the primary efficacy analysis, stratified by age group, [redacted].

5 INTERIM ANALYSIS

This study will utilize an independent Data Safety Monitoring Committee (DSMC) with the primary responsibility to monitor both blinded and unblinded information relating to subject safety during the study. A DSMC interim review meeting will be triggered approximately every 3 months following the commencement of the first patient being randomized and dosed with study medication (tideglusib or placebo).

All study staff including the Sponsor will remain blinded until the end of the study.

Full details of the objectives, timing, analyses and the role of the DSMC will be provided in the DSMC charter and DSMC SAP Analysis Plan.

6 ANALYSIS PLAN

6.1 General

Summary statistics for continuous variables will consist of number of non-missing observations (n), mean, standard deviation (SD), minimum, median, maximum, 25th and 75th percentiles unless specified otherwise.

For categorical variables the number and percentage of subjects in each category will be presented, based on the number of subjects in the analysis set, unless otherwise specified.

All statistical tests will be performed using a two-tailed 5% overall significance level, unless otherwise stated. The null hypothesis at all times will be that the treatments are equivalent. All comparisons between the treatments will be reported with 95% confidence intervals for the difference.

6.2 Blinded Data Review Meeting

The Sponsor will convene a blinded data review meeting (BDRM) after the data has been cleaned and before the study is unblinded.

The BDRM will make decisions that will include, but will not be limited to:

- the determination of whether protocol violations are 'important', or not a protocol violation at all;
- the allocation of subjects to analysis sets;
- changes required to the SAP.

If required, after the BDRM and prior to database lock, a SAP amendment will be issued.

6.3 General Derivations

This section provides details of general derivations. Derivations specific to the parameter of interest are detailed within the specific SAP section:

- **Randomized study treatment**

Where mentioned in this SAP, "Randomized study treatment" will relate to double-blind study treatment (either tideglusib or placebo). In the case of the SAF, this means the treatment that the subject received rather than the treatment that the subject was randomized to.

The first administration of randomized study medication is defined as the first dose of study medication after the date/time of randomization.

- **Phase**

Phase is defined as a subject's treatment period where:

  - Screening Phase: relates to the period ending the day of Visit [redacted] study medication will be dispensed at Visit [redacted] and dosing commences the next day.
  - [redacted] Run-In Phase: relates to the period starting the day after Visit [redacted] and ending immediately before the date/time of randomization at Visit [redacted]
  - Double-blind Treatment Phase: relates to the period starting immediately after the date/time of randomization and ending the day of the EOT visit, Visit [redacted]
  - Post-Treatment Phase: relates to the period starting from the day after the EOT visit, Visit [redacted].

If a visit is split across 2 days, completion of the full study visit is when all assessments on the 2nd day have been completed. Hence the end date of the visit should be used in phase derivations where applicable.

- **Definition of study day**

Throughout this SAP any references to "Visit XX" refer to the pre-specified visits defined in the protocol. Data listings will additionally present study day. Study day will be calculated relative to the first date of administration of randomized study treatment (Study Day 1). There is no Study Day 0. The day before Study Day 1 is Study Day -1.

- **Definition of baseline and run-in baseline**

Baseline is defined as the last non-missing value or mean values, where assessments are made in triplicate such as for ECGs, prior to the subject receiving randomized study treatment. Scheduled and unscheduled assessments will be considered.

Run-in baseline is defined as the last non-missing value or mean values, where assessments are made in triplicate such as for ECGs, prior to the subject receiving any study treatment, including [redacted]. Scheduled and unscheduled assessments will be considered. This baseline will be used to assess the effect of the [redacted] treatment taken during the [redacted] run-in period for selected safety data. For ECG, where assessments are performed in triplicate more than once at a visit, the relevant time-matched baselines will be used for the analysis of data, where applicable.

- **Definition of End of Treatment (EOT)**

End of treatment is defined as Visit [redacted] (Week [redacted] or the last visit in the study for subjects lost to follow-up or those who were withdrawn from the study by investigator.

- **Incomplete dates**

Every effort will be made to minimise incomplete dates. For calculation purposes, incomplete dates will be completed using worst case. Further details are detailed in the relevant sections as required.

- **Non-numeric values**

In the case where a variable is recorded as “>x”, “≥x”, “<x” or “≤x”, then for analysis purposes a value of x will be taken. Where a range of values is quoted the midpoint of the range will be taken. For example, if a laboratory safety parameter is reported as being below the limit of quantification or < x, the value of the limit will be used in the calculation of summary statistics. The recorded value will be reported in listings.

- **Methods for handling missing item responses in the Congenital DM1 Rating Scale (CDM1-RS)**

For this measure, either clinician-completed, caregiver-completed or independent rater-completed, one or more individual items may have a missing response at a given time point while the remaining items have been completed. Without imputation this would result in a missing total score. In order to avoid this scenario, each individual missing item will be replaced by the mean of that item alone from subjects within the same treatment group at that time point. The total score will then be computed using these values. This will be done before any other missing data methods are applied.

- **Methods for handling withdrawals and missing data**

For the primary and key secondary estimands, missing data will be accounted for using a direct likelihood approach (MMRM) and multiple imputation (MI) will be used in sensitivity analyses. MMRM models handle missing data by assuming it is missing at random (MAR), so that mean changes from baseline are estimated assuming that subjects with missing data will perform in the same way as subjects with the same covariate values and the same observed data so far. MI uses data from those subjects with similar characteristics to those with the missing data to impute a value for each subject and produces an imputed dataset. This process is repeated multiple times. For

each imputed dataset, an analysis is completed and the estimates from each analysis are combined to provide the estimates for the endpoint.

Some exploratory endpoints utilize ANCOVA models, with a subject with missing data at baseline and/or Visit [redacted] being excluded from the analysis. Missing data for these models will be assumed to be missing completely at random (MCAR). Analyses will also be done to explore the effect of missing not at random (MNAR) scenarios on the results.

Prior to database lock, the data will be reviewed to assess whether any subjects were unable to complete an assessment due to "extreme" factors affecting their ability to complete an assessment e.g. due to a serious deterioration in their physical state. In the event that such subjects are identified, additional appropriate methods for handling this specific type of missing data may be used.

Further details are given in section 6.13 below.

- **Visit Windows**

All subjects are expected to have an in-clinic visit for all pre-specified timepoints (except for the collection of AE and concomitant medication data at Visits [redacted] and [redacted] Should an in-clinic visit be missed a telehealth visit can replace it using the following rules:

Telehealth visits falling within [redacted] days of the expected date of a post-randomization in-clinic visit up to Visit [redacted] and [redacted] days of the expected date of Visits [redacted] and [redacted] where the expected date is calculated from the randomization visit (Visit [redacted].

6.4 Analysis Sets

The following analysis set definitions will be used:

The Enrolled Set includes all subjects for whom informed consent has been provided and assent provided (if applicable), irrespective of whether they received any study treatment.

The **Run-In Set (RIS)** consists of all subjects who received at [redacted] The Run-In set is used to assess the effect of the [redacted] period for selected efficacy data only.

The **Safety Analysis Set (SAF)** consists of all randomized subjects who received at least one dose of study treatment. Subjects will be presented by treatment actually received. The SAF is the primary population for the analysis of safety data.

The Intent-To-Treat (ITT) set consists of all randomized subjects with at least one post-baseline efficacy assessment. Subjects will be presented in the treatment group they were randomly assigned to. The ITT is the primary population for the analysis of efficacy data.

The Full Analysis Set (FAS) will include all randomised subjects in the SAF for whom any efficacy data were collected post randomisation. Subjects will be included in the analysis set according to the treatment they actually received.

The Pharmacokinetic Analysis Set (PKAS) consists of all subjects in the ITT who receive tideglusib and have at least one evaluable post-dose PK measurement (even if below the limit of quantification). Evaluable PK measurements are defined as those

with dosing information prior to sample collection and sample collection information available. Subjects will be analyzed according to the treatment actually received.

The Per-Protocol Set (PPS) will include all subjects in the FAS who did not have any important protocol violations. Definitions of important protocol violations will be included in the SAP prior to data lock and unblinding. Subjects excluded from the PPS will be determined prior to data unblinding. If no subjects are excluded from the PPS, the PPS will not be required. The PPS will be used for secondary analyses of efficacy data.

All protocol deviations will be assessed and documented on a case-by-case basis prior to the database lock. Important protocol deviations are a subset of protocol deviations that may significantly impact the completeness, accuracy, and/or reliability of the study data or that may significantly affect a subject's rights, safety, or well-being. Important protocol deviations and their impact on the analysis sets for this study are specified in the Protocol Deviations Criteria (PDC) form. Only important protocol deviations will be included in the statistical outputs.

The definitions for the Enrolled, PKAS, RIS, SAF, ITT and FAS analysis sets are sufficient to determine the subjects included within these analysis sets and so do not require listing and manual review for approval.

6.5 Data presentations

The TFLs will present treatment group (i.e. tideglusib or placebo).

TFLs on the PKAS, SAF and FAS analysis sets will be presented according to the treatment actually received (i.e. tideglusib or placebo), regardless of the randomization.

TFLs on the Enrolled, ITT and PPS analysis sets will be presented according to the randomized treatment group.

The data will be summarised in tabular form by treatment group apart from disposition of subjects, protocol deviations and background and demographic data which will be summarised by treatment group and overall subjects.

Only scheduled post-baseline laboratory, vital signs and ECG values will be tabulated, post-baseline repeat/unscheduled assessments will be listed only; all clinically significant values will be noted.

Individual subject assignments to each analysis set will be summarized using the Enrolled set. The primary and key secondary efficacy endpoints will be summarised using the ITT and the PPS. All other secondary endpoints will be summarised using the ITT. PK concentration data will be summarised using the PKAS set and all other data will be summarized using the SAF set. Additionally, [redacted] run-in data for efficacy and adverse events will be summarised for the RIS set.

Eligibility, completion/withdrawal, consent/assent, analysis set, protocol deviations, demographic and other background listings in section 6.8 and visit dates will be based on the enrolled set, PK listings will be based on the PKAS set, efficacy listings, [redacted] run-in treatment, study drug dispensing and return, study drug daily dosing and meal intake data, all adverse events and adverse events leading to death will be based on the RIS set and all other listings will be based on the SAF set. Listings based on

the RIS will presented according to the randomized treatment for the efficacy listings and actual treatment received otherwise.

Graphical presentations of the data will also be provided where appropriate. Unless otherwise stated, only data from Visit [redacted] onwards will be presented graphically.

6.6 Disposition of subjects

The number and percentage of all subjects will be presented overall and by site for the following:

- Enrolled set
- Screening failures including inclusion criteria failed and exclusion criteria met.
- RIS set
- Run-in failures.
- Subjects randomized.
- SAF, ITT, FAS, PPS and PKAS sets.
- Subjects who completed the study.
- Subjects who prematurely discontinued from the study including reason for withdrawal.

The percentage of screening failures will be based on the enrolled set and the percentage of run-in failures will be based on the RIS. Otherwise percentages will be based on the number of subjects randomized.

In addition, total duration, derived as (date of study completion or the date of early study withdrawal - date of first administration of study treatment including [redacted] +1, and double-blind treatment duration derived as (date of study completion or the date of early study withdrawal - date of first administration of randomized study treatment will be summarized.

Screening failures are defined as subjects who do not take at least [redacted] and who discontinue the study. Run-in failures are defined as subjects who take a [redacted].

Subjects who prematurely discontinued from the study due to a reason relating to COVID-19 are identified as either of the following:

- Reason for withdrawal of 'Other' with a corresponding free text reason containing a text version of 'COVID-19'.
- Reason for withdrawal of 'Adverse Event' with a corresponding COVID-19 related AE reported on the AE eCRF page.

The number and percentage of subjects who prematurely discontinued due to either of these definitions will be presented as a reason for withdrawal of 'COVID-19'.

Eligibility for each of the analysis sets and reasons for exclusion will be listed. Eligibility criteria, informed consent and assent, study completion/withdrawal and Visit dates data will be listed.

6.7 Protocol Deviations

Prior to database lock, AMO Pharma will review the individual deviations to confirm that important protocol deviations have been captured correctly (as defined in section 6.4 above) and to agree impact to analysis sets as indicated in the Protocol Deviation Criteria (PDC) form.

Only important protocol deviations, as specified in the PDC form, will be included in the statistical outputs.

Details of important protocol deviations (date, deviation category and specific details) and subject eligibility will be listed.

The number and percentage of subjects with at least one important protocol deviation will be summarised for each deviation category.

6.8 Background and Demographic Characteristics

6.8.1 Demography

Demographic characteristics (age, sex, ethnic origin and race), body measurements (height, weight and BMI), dominant arm (left or right) and number of CTG repeats (if available) collected at Screening will be summarised overall and by site. In addition, date of birth (DOB) will be listed. Subjects enrolled in New Zealand will have a DOB captured as 01 MMM YYYY.

Age is calculated in years from the date of the screening visit.

All subject demographic data including informed consent/assent will be listed.

6.8.2 Medical History

Medical and surgical history events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) dictionary version applicable at eCRF go-live. The version used will be indicated in the data summaries and listings. Previous and ongoing conditions will be presented separately. Previous conditions are defined as those that started and ended prior to the first administration of study treatment, including [redacted]. All other conditions will be assumed ongoing.

The number and percentage of subjects will be presented by system organ class (SOC), and preferred term (PT), where SOC and PT will be presented in decreasing frequency of the total number of subjects with medical or surgical history events. In summary tables, subjects with medical or surgical history in the same SOC or having the same preferred term recorded multiple times will be counted only once for that SOC and PT.

All events will be listed.

6.8.3 Symptoms of Disease

Each symptom of disease under study is captured as free text with a start date and severity (mild/moderate/severe). All symptoms of disease under study data will be coded using the MedDRA dictionary applicable at eCRF go-live with the MedDRA version used being indicated.

All symptoms of disease data will be listed. The MedDRA version used will be indicated.

6.8.4 Caregiver DM1 Status

Caregiver DM1 status data will be listed.

6.8.5 Other Data

[redacted] sampling data, test telehealth session data and telehealth type data will be listed.

6.9 Prior, Concomitant and Follow-up Period Medications

Medications will be coded using the latest World Health Organization (WHO) Drug dictionary version applicable at eCRF go-live. The version used will be indicated in the data summaries and listings.

Prior medications are defined as those that started and ended prior to the first administration of randomized study treatment. Medications that ended on or before a subject's last dose of randomized study medication and are either ongoing at the first administration of randomized study treatment or started after the time of first administration will be deemed to be concomitant medications. If medication dates are incomplete and it is not clear whether the medication was concomitant, it is assumed to be concomitant. Follow-up period medications are defined as those that started after a subject's last dose of randomized study medication.

Prior, concomitant and follow-up period medications will be summarized separately.

The number and percentage of subjects taking medications will be summarised by medication class and standardised medication name, where medication class and standardised medication name will be presented in decreasing frequency of the total number of subjects with medications. In summary tables, subjects taking multiple medications in the same medication class or having the same standardised medication recorded multiple times in the study will be counted only once for that specific medication class and standardised medication name.

Medication data will be listed, where medications will be flagged as prior, concomitant or follow-up period medications.

6.10 Non-Pharmacological Treatment Therapies

Non-Pharmacological treatment therapy data will be listed.

6.11 Administration of Study treatment and Exposure

6.11.1 [redacted] Run-In Phase

During the [redacted] run-in phase, [redacted] will be administered once each day at approximately the same time each day, by oral route or gastrostomy tube. Subjects will receive [redacted] which will be weight adjusted.

[redacted] treatment administration percentage compliance is derived as follows:

$$\left(\frac{Total\ amount\ of\ [redacted]\ treatment\ taken}{Total\ amount\ of\ [redacted]\ treatment\ expected\ to\ be\ taken}\right) x100$$

Where treatment dispensing commences from Visit [redacted] and the total amount (mg) of [redacted] treatment taken is given by:

[redacted] *sum of* [redacted] *dispensed – sum of* [redacted] *returned) +* [redacted] *(sum of* [redacted] *dispensed – sum of* [redacted] *returned).*

The total amount (mg) of [redacted] treatment expected to be taken is derived as follows:

Treatment period x 1000, where the treatment period is given by: (Visit [redacted] *or withdrawal date if prior to Visit* [redacted] *– Visit* [redacted] *date.*

Any percentage compliance derived to be in excess of 100% will be set to 100%.

Percentage compliance with administration of run-in treatment will be summarised.

6.11.2 Double-Blind Treatment Phase

During the double-blind treatment phase study treatment (tideglusib or matched placebo) will be administered once each day at approximately the same time each day, by oral route or gastrostomy tube. Dosing will be weight-adjusted at 400 mg, 600 mg, or 1000 mg dose levels, with each subject starting at a weight-adjusted 400 mg dose level for [redacted] weeks, then [redacted] to a weight-adjusted 600 mg dose level for the next [redacted] weeks until they reach the final dose level of weight-adjusted 1000 mg. [redacted] to [redacted] is permitted during the double-blind treatment period. Subjects [redacted] will be assigned to the actual treatment group they received most of the time during the study.

Dosing will be weight adjusted at tideglusib 400 mg, 600 mg or 1000 mg dose levels, using either a [redacted], a [redacted] or both a [redacted], respectively. For blinding purposes, each active dose has a corresponding matching placebo presentation. Regardless of the dose a subject has been assigned to, study medication will always be prepared using [redacted]. This can be a [redacted], an [redacted], or an [redacted] depending on which treatment group the subject has been assigned to and which dose they are taking, e.g. [redacted]

Randomized treatment administration percentage compliance is derived as follows:

$$\left(\frac{Total\ amount\ of\ randomized\ treatment\ taken}{Total\ amount\ of\ randomized\ treatment\ expected\ to\ be\ taken}\right) x100$$

Where randomized treatment dispensing commences from Visit [redacted] and the total amount (mg) of randomized treatment taken is given by:

[redacted] *(sum of* [redacted] *dispensed – sum of* [redacted] *returned) +* [redacted] *(sum of* [redacted] *dispensed – sum of* [redacted] *returned).*

The total amount (mg) of randomized treatment expected to be taken is derived as follows:

1. [redacted] step 1 (Visit [redacted] to Visit [redacted]): Treatment period [redacted] where the treatment period is given by: (Visit [redacted] or withdrawal date if prior to Visit [redacted]) – Visit [redacted] date.
2. [redacted] step 2 (Visit [redacted] to Visit [redacted]): Treatment period x [redacted] where the treatment period is given by: (Visit [redacted] or withdrawal date if prior to Visit [redacted] – Visit [redacted] date.
3. Maintenance dose level for subjects [redacted] (Visit [redacted] to Visit [redacted]): Treatment period x [redacted] where the treatment period is given by:

   (Visit [redacted] or withdrawal date if prior to Visit [redacted] – Visit [redacted] date.
4. Maintenance dose level for subjects with [redacted] (Visit [redacted] to Visit [redacted]): [redacted] x (treatment period prior to [redacted]) + [redacted] x (treatment period after [redacted], where the treatment period prior to [redacted] is given by:

   Date subject stopped taking higher dose of active treatment – Visit [redacted] date and the treatment period after [redacted] given by:

   (Visit [redacted] or withdrawal date if prior to Visit [redacted] – date (taken from diary card) subject started taking lower dose of randomized treatment + 1.

Any percentage compliance derived to be in excess of 100% will be set to 100%.

Percentage compliance with administration of randomized treatment will be summarised.

The following randomized treatment exposure summaries will also be provided: total amount (mg) of randomized treatment taken and the number of days of exposure to the randomized treatment derived from the subject diary card data as follows:

(Date of last dose of higher dose of randomized treatment – date of first dose of higher dose of randomized treatment + 1) + (Date of last dose of lower dose of randomized treatment – date of first dose of lower dose of randomized treatment + 1), where the lower dose duration is only applicable to subjects who are down-titrated.

In addition, the following randomized treatment exposure summaries will be provided:

- the number of days of exposure to [redacted] step 1 [redacted] tideglusib/matched placebo defined as:

  Date of last dose of [redacted] randomized treatment – date of first dose of [redacted] randomized treatment + 1.
- the number of days of exposure to [redacted] step 2 [redacted] tideglusib/matched placebo defined as:

  Date of last dose of [redacted] randomized treatment – date of first dose of [redacted] randomized treatment + 1. Any [redacted] tideglusib/matched placebo received during [redacted] should be excluded.
- the number of days of exposure to [redacted] tideglusib/matched placebo for all subjects and separately for subjects who are [redacted] from [redacted] defined as:

  Date of last dose of [redacted] randomized treatment – date of first dose of [redacted] randomized treatment + 1.
- the number of days of exposure to [redacted] tideglusib/matched placebo for subjects who are down-titrated from [redacted] tideglusib/matched defined as:

Date of last dose of [redacted] randomized treatment– date of first dose of [redacted] randomized treatment + 1. Any [redacted] tideglusib/matched placebo received during [redacted] should be excluded.

- the number of subjects that were [redacted]

The above derived variables and study drug dispensing, study drug return, study drug [redacted] and subject diary card information will be listed.

6.12 COVID-19 Impact to study

Due to the global COVID-19 (Coronavirus Disease 2019) pandemic and the potential for resurgences of COVID-19, for the purposes of this protocol, 'in-clinic' visits, when pertaining to Visits [redacted]-Visit [redacted] (Weeks [redacted]), may be conducted via a combination of telehealth and home healthcare providers for COVID-19 related reasons e.g. institutional restrictions, travel restrictions, quarantine or shielding prevent subjects from being seen in-person at the investigator site.

Subjects who have completed Screening (Visit [redacted] but are unable to attend Run-In (Visit [redacted] in-person due to COVID-19 related reasons will be allowed the opportunity to rescreen for the study if study enrolment is still ongoing once they are able to attend the clinic again in person. These subjects will have a new screening number and will be treated as a separate subject in the tables and listings. Subjects who have attended the Run-In (Visit [redacted] and are unable to attend the Baseline (Visit [redacted] in-person due to COVID-19 related reasons will be discontinued and replaced.

All subjects are expected to have an in-clinic visit for all pre-specified timepoints (except for the collection of AE and concomitant medication data at Visits [redacted] and [redacted] Should an in-clinic visit be missed the corresponding telehealth visit and/or home health care visit can replace it. Additionally, there will be 3 occasions in the study (Visit [redacted], Visit [redacted] and Visit [redacted] where dual telehealth and in-clinic assessments will be conducted. To evaluate the consistency of telehealth data and in-clinic data for the CDM1-RS and CGI rating scales and to minimize the potential bias, Visits [redacted] and [redacted] are performed using both in-clinic and telehealth data collection approaches. Sensitivity analyses to evaluate the consistency of these two methods of data collection is provided in sections 6.13.1 and 6.13.2.

It is noted that particular attention may need to be given to the impact on key efficacy or safety assessments as a result of the COVID-19 pandemic. Consideration of the different areas of impact has been made as described below:

- Important protocol deviations related to COVID-19 will be recorded under the deviation category of 'Other' with a free text reason indicating the deviation was due to COVID-19. The assigned categories will be presented in the protocol deviation listing and table summary of important deviations as described in section 6.7.

- Subjects who prematurely discontinue the study due to COVID-19 will be identified as described in section 6.6 and will be presented as a reason for withdrawal of 'COVID-19'.

- Missing efficacy data for the primary and key secondary estimands will be subject to the imputation and analysis methods described in sections 6.13.1 and 6.13.2.

In order to provide supportive information for interpretation of the key efficacy and safety analyses, a summary of the number and percentage of subjects with missing in-clinic (telehealth assessment only being available) and completely missing assessments (no in-clinic or telehealth assessment being available) at each scheduled visit will be presented by treatment group and overall for the following selected data on the ITT analysis set or safety analysis set (safety data) as appropriate:

- Primary and secondary efficacy endpoints (specifically in-clinic and telehealth Clinician Completed CDM1-RS, clinician in-clinic and telehealth CGI-I, Top 3 Caregiver concerns, Caregiver Completed CDM1-RS, 10-meter walk-run test)
- Clinical chemistry and complete blood counts
- ECGs

A summary of missing items for the clinician in-clinic and telehealth CDM1-RS and caregiver CDM1-RS will also be produced.

If these summaries suggest that further investigation is beneficial, additional summaries and/or analyses may be added post hoc to address missing data or assessments outside of protocol defined windows. This may include but is not limited to: additional sensitivity analyses of the efficacy endpoints, additional summaries of selected safety data for overall worst case post-baseline values, application of visit windows to group post-baseline data across multiple visits as appropriate.

6.13 Efficacy Evaluation

Unless specified otherwise, all listings will be based on the RIS set.

All efficacy analyses and summaries will be primarily performed on the ITT analysis set. Summaries of run-in efficacy data will also be performed on the RIS set. The analysis for the primary and key secondary estimands will be repeated on the FAS and the PPS sets. The sensitivity analyses of the primary and the key secondary estimands will be based on the ITT set only. All other efficacy analyses will be based on the ITT set only.

All hypothesis tests comparing tideglusib and placebo will be two-sided and conducted at the 5% significance level, unless otherwise specified.

6.13.1 Clinician-Completed Congenital DM1 Rating Scale (CDM1-RS)

The CDM1-RS is completed by the clinician to score the symptom severity of the items that are clinically relevant in Congenital DM1. The severity of the clinician's concern in each domain is scored by using a 5-point Likert Scale. Scores range from 0 = Not present to 4 = Very severe. The clinician is asked to rate the severity of each symptom, using a time frame of the past week including the day of the assessment for reference.

The CDM1-RS is administered at Visits [redacted] and [redacted] At Visits [redacted] and 11, the CDM1-RS will be rated twice for consistency evaluation purposes. One rating will be from an interview completed in-clinic and an another from an interview completed via telehealth within the [redacted] period before the in-clinic visit.

The CDM1-RS total score will be derived at each visit as the total score based on 11 domain/item ratings, with a higher total score indicating worse symptom severity.

6.13.1.1 Primary Estimand and Supplementary Analysis

The estimand for the primary efficacy endpoint is the difference between group adjusted means in the changes from baseline in CDM1-RS total score at EOT.

If a subject's in-clinic visit is unavailable, the data will be imputed with the corresponding telehealth assessment. The use of telehealth assessments is thought unlikely to be related to outcome, however, it will be assessed for balance between the treatment groups and sensitivity analyses may be conducted in the event of imbalance.

The null hypothesis being tested for the primary endpoint is that there is no difference in mean change in CDM1-RS total score from baseline to EOT between the tideglusib and placebo treatment groups:

$$H_0: Mean\ Change_{tideglusib} = Mean\ Change_{placebo}$$

The alternative hypothesis is that there is a difference in either direction in the mean change in CDM1-RS total score from baseline to EOT between the tideglusib and placebo treatment groups:

$$H_1: Mean\ Change_{tideglusib} \neq Mean\ Change_{placebo}$$

The analysis will include data from Visit [redacted] (randomisation) to Visit [redacted] (EOT) inclusive. Data from other visits will be included in the summaries and listings.

Primary Analysis

The primary estimand will be analysed using a mixed model for repeated measures (MMRM).

The treatment difference between tideglusib and placebo at Visit [redacted] (EOT) will be estimated as the simple contrast in the treatment effect. The standard error (SE), p-value and 2-sided 95% confidence interval (CI), based on the difference in the Least Squares (LS) means between the two treatment groups will be presented. The LS Means, SE and 95% CI will additionally be presented for each treatment by visit.

The MMRM model will be implemented using restricted maximum likelihood (REML). The Kenward-Roger approximation will be used to estimate the denominator degrees of freedom and standard errors

Initially an unstructured covariance matrix will be investigated. If this analysis fails to converge, the following alternative structures will be tested in the following order until convergence is achieved: Toeplitz, AR(1) and compound symmetry. If a covariance matrix other than unstructured is selected, the estimated standard errors and variance-covariance matrix of the fixed-effects parameters will be computed using a robust "sandwich" estimator (i.e., by specifying the EMPIRICAL option on the PROC MIXED statement).

The model will contain treatment, age group, visit and treatment by visit interaction as fixed effects, baseline CDM1-RS as a covariate.

If the baseline CDM1-RS and/or age group is not significant at the 15% level (i.e., p>=0.15) then a secondary analysis will be conducted with the term(s) removed from the model. Similarly if the p-value for the treatment by visit interaction is not significant at the 15% level, that term will be removed in the secondary analysis and the LS

Means and treatment differences under the assumption of a constant treatment effect over time will be estimated.

Normality of the data will be investigated through plots of the residuals. Non-parametric analyses may additionally be performed if the data are found to be sufficiently non-normal. This will include non-parametric analysis of covariance (ANCOVA) (Koch G.G. et al, 1998 and Zink, R. and Koch, G.G., 2012) of the change from baseline to Visit [redacted] with treatment, baseline CDM1-RS and age group as covariates.

Any data collected after study discontinuation will not contribute to the treatment effect estimate. Only data up to study discontinuation will be used.

Missing Data

Missing data due to early study discontinuations or missed visits are assumed to be missing at random (MAR) and observed data only will be used for the primary analysis. The MMRM accounts for these missing data using a direct likelihood approach.

Sensitivity Analysis

Several sensitivity analyses will be applied to this endpoint:

Sensitivity analysis 1 – Multiple Imputation: Missing at Random (MAR)

At a visit, if a subject has no in-clinic or telehealth assessment data the visit is assumed to be MAR. This assumes that subjects with missing data follow the same trajectory as other subjects in their respective treatment arm that have complete data.

- Intermittent missing data will first be imputed using the MCMC method, appropriate for non-monotonic missing data.
- Data missing after subjects discontinue the study early will be imputed using a regression, appropriate for monotone missingness. At each time point, missing data will be assumed to follow a distribution similar to scores for subjects who are still in the study and randomized to the same treatment group.

The MI method (as described in Section 6.3) will be employed with the following assumptions/steps:

- Covariates that may be used in the MI will include treatment group, age group and baseline CDM1-RS.
- 20 imputations will be completed.
- Seed number and covariates used for imputation will be agreed prior to data base lock and detailed in the TFL outputs.

The MMRM will be applied to each of the multiply imputed data sets, as described above for the primary estimand. The estimates from these models will then be combined using Ruben's (1987) rule and analysed to provide the endpoint estimates and summarized.

Sensitivity analysis 2 – Single Imputation: WOCF

At a visit, if a subject has no in-clinic or telehealth assessment data the visit is assumed to be MAR. This assumes that the trajectory of subjects who discontinue the study follows that of their worst observation. Worst Observation Carried Forward (WOCF)

will be used and the MMRM as described above for the primary estimand will be applied to these data.

Sensitivity analysis 3 – Multiple Imputation: In-clinic visits only

At a visit, if a subject has a telehealth assessment and no in-clinic assessment the visit is assumed to be MAR. Modelling in-clinic visits only (sensitivity analysis 3) is deemed a more valuable analysis than modelling telehealth visits only (sensitivity analysis 4).

- Intermittent missing data for in-clinic visits will first be imputed using the MCMC method, appropriate for non-monotonic missing data.
- Data missing after subjects discontinue the study early will be imputed using a regression, appropriate for monotone missingness. At each time point, missing data will be assumed to follow a distribution similar to scores for subjects who are still in the study and randomized to the same treatment group.

The MI procedure will be employed as described for sensitivity analysis 1. The MMRM will be applied to the multiply imputed data sets, as described above for the primary estimand. The estimates from these models will then be combined using Ruben's (1987) rule and analysed to provide the endpoint estimates and summarized.

Sensitivity analysis 4 – Multiple Imputation: Telehealth visits only

At a visit, if a subject has an in-clinic assessment and no telehealth assessment the visit is assumed to be MAR.

- Intermittent missing data for telehealth visits will first be imputed using the MCMC method, appropriate for non-monotonic missing data.
- Data missing after subjects discontinue the study early will be imputed using a regression, appropriate for monotone missingness. At each time point, missing data will be assumed to follow a distribution similar to scores for subjects who are still in the study and randomized to the same treatment group.

The MI procedure will be employed as described for sensitivity analysis 1. The MMRM will be applied to the multiply imputed data sets, as described above for the primary estimand. The estimates from these models will then be combined using Ruben's (1987) rule and analysed to provide the endpoint estimates and summarized.

Sensitivity analysis 5 – Completers

The MMRM will be applied to all subjects who have completed the study to EOT.

Sensitivity Analysis 6 – Rank Analysis

The change from baseline in CDM1-RS total score at EOT will be ranked with any tied observations being assigned an average rank. The sum of the ranks for each treatment will be calculated. The non-parametric Wilcoxon rank-sum test will be utilised at a 2-sided 5% level of significance to test the null hypothesis that the distribution of post randomization CDM1-RS total score values for tideglusib and placebo is equal against the alternative hypothesis that the distribution of post-randomization CDM1-RS total score values for tideglusib and placebo are not equal. The Hodges-Lehmann estimate of the difference between the two treatments and 95% confidence interval will also be calculated.

Sensitivity analysis 7 – Multiple Imputation: MNAR control-based pattern

For this analysis, in-clinic and telehealth data will be used as for the primary analysis.

- Intermittent missing data for in-clinic visits will first be imputed using the FCS method, appropriate for non-monotonic missing data.
- Data missing after subjects discontinue the study early will be imputed using a regression, appropriate for monotone missingness. At each time point, missing data will be assumed to follow a distribution similar to scores for subjects who are still in the study and randomized to placebo.

The MI procedure will be employed as described for sensitivity analysis 1. The MMRM will be applied to the multiply imputed data sets, as described above for the primary estimand. The estimates from these models will then be combined using Ruben's (1987) rule and analysed to provide the endpoint estimates and summarized.

Sensitivity analysis 8 – Multiple Imputation: MNAR delta-adjustment pattern

For this analysis, in-clinic and telehealth data will be used as for the primary analysis.

- Intermittent missing data for in-clinic visits will first be imputed using the FCS method, appropriate for non-monotonic missing data.
- Data missing after subjects discontinue the study early will be imputed using a regression, appropriate for monotone missingness. At each time point, missing data will be assumed to follow a distribution similar to scores for subjects who are still in the study and randomized to the same treatment group. However, a fixed 5% worsening (delta-adjustment) to the outcome in the tideglusib group will then be applied.
- In the event of a statistically significant treatment effect, several delta-adjustments will be performed until non-significance occurs. This will be used to identify how much worse a response would need to have occurred in the subjects with missing data in order to have overturned the significant result (a tipping-point analysis).

The MI procedure will be employed as described for sensitivity analysis 1. The MMRM will be applied to the multiply imputed data sets, as described above for the primary estimand. The estimates from these models will then be combined using Ruben's (1987) rule and analysed to provide the endpoint estimates and summarized.

<u>Supplementary Analysis</u>

- The primary efficacy analysis will be repeated on the FAS and PPS.
- The proportion of subjects who have a 10% or greater improvement over baseline in the CDM1-RS total score will be summarised and compared using a Fishers Exact test. Similar analyses using 15% and 20% or greater improvements will also be undertaken.
- Subgroup analyses will be performed by key baseline characteristics. The balance between treatment groups will be assessed and if any are evaluated as not being consistent a respective covariate will be included in the MMRM. For each subgroup, the main effect, treatment-by-subgroup interaction and treatment by visit by subgroup interaction terms will be added to the model. Any interactions that are statistically significant at the 15% level for the end of treatment visit will have their nature described. These models will be used to estimate treatment comparisons within the subgroups that correspond with the

sub-grouping factor. The results from the analysis are to be presented as described above in tables and/or figures. Subgroups of interest include but are not limited to:

- Race (White, Non-White)
- Sex (Male, Female)
- Age group ([redacted])
- Improvement during the [redacted] run-in period (Yes, No). Improvement in this case is defined as a 2-point or greater improvement in the total score.

- To identify items which are primarily driving an improvement in CDM1-RS total score, the 11 individual items will each be analysed in an analogous method to the primary estimand above. A forest plot with all 11-items on the y-axis, showing the treatment effect will additionally be provided.
- The occurrence and potential imbalance of intercurrent events (e.g. concomitant medication changes, protocol non-compliance), deaths and COVID-19 impacts between treatment groups will be reviewed after unblinding the randomization.
    - If an imbalance is detected in the intercurrent events, further summaries and/or analyses may be undertaken to assess the impact of this.
    - Deaths are not anticipated during the study. If an imbalance in deaths is determined, a joint rank analysis (based on the change in CDM1-RS from baseline and time to death) may be utilized.

Summaries and Figures

Summary statistics of observed values and change from baseline will be presented by visit and treatment for the in-clinic and telehealth visits separately. For subjects who have both an in-clinic and telehealth assessment at a visit, change from baseline in CDM1-RS will be calculated for both methods and summarised separately.

The observed CDM1-RS total score will be presented over time using by subject line plots by treatment group for the in-clinic and telehealth visits separately. The plots will be presented together but on two separate panels.

Additionally, histograms of the change from baseline at each visit will be presented by treatment group. Empirical cumulative distribution functions (CDFs) of the change from baseline will be presented by treatment group at Visit [redacted]

The adjusted LS mean changes from baseline and SEs over time will be presented via line plots as estimated by the primary estimand MMRM model by treatment group.

For the 3 occasions in the study (Visit [redacted] and [redacted]) where dual telehealth and in-clinic assessments are conducted, summary statistics of the difference between the methods (in-clinic – telehealth) will be calculated and summarized using descriptive statistics over time. The intra-rater correlation will also be presented for each visit. A scatterplot for each visit will compare the observed value for the two methods for each subject, with in-clinic assessments being presented on the y-axis and telehealth on the x-axis. A linear regression line will be included. A line plot will present the mean changes from baseline to each visit, with in-clinic and telehealth assessments being presented on the same plot. Corresponding individual profile plots will also be presented.

A forest plot will display the results of the primary estimand model, the interaction p-values for each subgroup, and the descriptive results within each subgroup level.

Data for in-clinic and telehealth assessments will be listed separately, with whether a telehealth assessment is within the pre-defined visit window for imputation as an in-clinic assessment being annotated.

Measurement Properties of the CDM1-RS

The measurement properties of the CDM1-RS will be addressed as follows:

- To assess whether individual items scores have undue impact on the total score, item-total Pearson correlations (and p-values) will be computed. This will be repeated for the change from baseline scores.
- To assess the sensitivity to change of the CDM1-RS, the total score and individual item scores will be correlated with the CGI-I and changes in the 10-metre walk/run time at preferred speed and fastest speed using Pearson correlations (and p-values). Note that it is not certain if the 10-metre walk/run time is itself sensitive to change in this study population.
- To assess the clinical meaningfulness of changes in the CDM1-RS total score, the change and percentage change from baseline in the CDM1-RS total score will be summarised by CGI-I value. The minimum clinically important difference is the change associated with "minimal improvement" on the CGI-I.

These analyses will use observed data from Visit [redacted] (and baseline where change from baseline is calculated) and performed on the ITT set overall (i.e., without reference to treatment group.) In addition, for those subjects that achieve "minimal improvement" on the CGI-I at any visit, the change and percentage change from baseline in the CDM1-RS total score from the first such visit will be listed and summarised.

6.13.2 Clinical Global Impression- Improvement Scale (CGI-I)

The clinician administered Clinical Global Impression-Improvement (CGI-I) rating scale permits a global evaluation of the subject's improvement over time. The CGI-I requires the clinician to rate how much the subject's illness has changed (improved, worsened or stayed the same) relative to a baseline state via a seven-point Likert type scale which ranges from 1 = very much improved to 7 = very much worse.

The CGI-I is administered at Visits [redacted] and [redacted] At Visits [redacted] and [redacted], the CGI-I will be rated twice for consistency evaluation purposes. One rating will be from an interview completed in-clinic and another from an interview completed via telehealth within the [redacted] before the in-clinic visit.

The CGI-I at Visit [redacted] is rated by the investigator against the subject's clinical presentation at Visit [redacted] This allows for an understanding of [redacted] during the [redacted] run-in period. For subsequent CGI-I ratings completed from Visit [redacted] onwards, the assessment is rated by the investigator against the subject's clinical presentation at Visit [redacted] A negative observed value indicates an improvement since baseline, whilst a positive observed value indicates a worsening.

The CGI-I may be assessed via telehealth instead of an in-clinic interview if required due to a COVID-19 related reason. The use of telehealth assessments is thought unlikely to be related to outcome, however, it will be assessed for balance between

the treatment groups and sensitivity analyses may be conducted in the event of imbalance.

6.13.2.1 Key Secondary Estimand and Supplementary Analyses

The estimand for the key secondary efficacy endpoint is the difference between treatments in the observed CGI-I score at EOT. Any data collected after study discontinuation will not contribute to the treatment effect estimation.

If a subject's in-clinic visit is unavailable, the data will be imputed with the corresponding telehealth assessment.

The null hypothesis being tested for the key secondary estimand is that there is no difference in CGI-I at EOT between the tideglusib and placebo treatment groups:

$$H_0: MeanCGI - I_{tideglusib} = MeanCGI - I_{placebo}$$

The alternative hypothesis is that there is a difference in either direction in the CGI-I at EOT between the tideglusib and placebo treatment groups:

$$H_1: MeanCGI - I_{tideglusib} \neq MeanCGI - I_{placebo}$$

The analysis will include data from Visit [redacted] (randomisation) to Visit [redacted] (EOT) inclusive. Data from other visits will be included in the summaries and listings.

Key Secondary Analysis

The key secondary estimand will be analysed using a MMRM.

The treatment difference between tideglusib and placebo at EOT will be estimated as the simple contrast in the treatment effect. The 2-sided 95% confidence interval (CI), based on the difference in the Least Squares (LS) means between the two treatment groups will be presented. The LS Means (and 95% CI) will additionally be presented for each treatment by Visit.

The MMRM model will be implemented using REML. The Kenward-Roger approximation will be used to estimate the denominator degrees of freedom and adjust standard errors.

Initially an unstructured covariance matrix will be investigated. If this analysis fails to converge, the following alternative structures will be tested in the following order until convergence is achieved: spatial spherical, spatial power, spatial Gaussian and compound symmetry. If a covariance matrix other than unstructured is selected, the estimated standard errors and variance-covariance matrix of the fixed-effects parameters will be computed using a robust "sandwich" estimator (*i.e.,* by specifying the EMPIRICAL option on the PROC MIXED statement).

The model will contain treatment, age group, visit and treatment by visit interaction as fixed effects, baseline CGI-I as a covariate.

If the baseline CGI-I and/or age group is not significant at the 15% level (i.e., p>=0.15) then a secondary analysis will be conducted with the term(s) removed from the model. Similarly if the p-value for the treatment by visit interaction is not significant at the 15% level, that term will be removed in the secondary analysis and the LS Means and treatment differences under the assumption of a constant treatment effect over time will be calculated.

Normality of the data will be investigated through plots of the residuals. Non-parametric analyses may additionally be performed if the data are found to be sufficiently non-normal. This may include non-parametric ANCOVA (Koch G.G. et al, 1998 and Zink, R. and Koch, G.G., 2012) of the CGI-I at Visit [redacted] with treatment, baseline CGI-I and age group as covariates. If the majority of subjects use three or fewer response categories, an approach suitable for categorical (binomial or multinomial) data may be used instead.

Any data collected after study discontinuation will not contribute to the treatment effect estimate. Only data up to study discontinuation will be used.

Missing Data

Missing data due to early study discontinuations or missed visits are assumed to be missing-at-random (MAR) and observed data only will be used for the primary analysis. The MMRM accounts for these missing data using a direct likelihood approach.

Sensitivity Analysis

Several sensitivity analyses will be applied to this endpoint:

Sensitivity analysis 1 – Multiple Imputation: MAR

At a visit, if a subject has no in-clinic or telehealth assessment data the visit is assumed to be MAR. This assumes that subjects with missing data follow the same trajectory as other subjects in their respective treatment arm that have complete data.

- Intermittent missing data will first be imputed using the MCMC method, appropriate for non-monotonic missing data.
- Data missing after subjects discontinue the study early will be imputed using a regression, appropriate for monotone missingness. At each time point, missing data will be assumed to follow a distribution similar to scores for subjects who are still in the study and randomized to the same treatment group.

The MI method (as described in Section 6.3) will be employed with the following assumptions/steps:

1. Covariates that may be used in the MI will include treatment group and age group.
2. 20 imputations will be completed.
3. Seed number and covariates used for imputation will be agreed prior to data base lock and detailed in the TFL outputs.

The MMRM will be applied to each of the multiply imputed data sets, as described above for the key secondary estimand. The estimates from these models will then be combined using Ruben's (1987) rule and analysed to provide the endpoint estimates and summarized.

Sensitivity analysis 2 – Single Imputation: WOCF

At a visit, if a subject has no in-clinic or telehealth assessment data the visit is assumed to be MAR. This assumes that the trajectory of subjects who discontinue the study follows that of their worst observation. WOCF will be used and the MMRM as described for the key secondary estimand and applied to these data.

Sensitivity analysis 3 – Multiple Imputation: In-clinic visits only

At a visit, if a subject has a telehealth assessment and no in-clinic assessment, the visit is assumed to be MAR. Modelling in-clinic visits only (sensitivity analysis 3) is deemed a more valuable analysis than modelling telehealth visits only (sensitivity analysis 4).

- Intermittent missing data for in-clinic visits will first be imputed using the MCMC method, appropriate for non-monotonic missing data.
- Data missing after subjects discontinue the study early will be imputed using a regression, appropriate for monotone missingness. At each time point, missing data will be assumed to follow a distribution similar to scores for subjects who are still in the study and randomized to the same treatment group.

The MI procedure will be employed as described for sensitivity analysis 1. The MMRM will be applied to the multiply imputed data sets, as described for the key secondary estimand. The estimates from these models will then be combined using Ruben's (1987) rule and analysed to provide the endpoint estimates and summarized.

Sensitivity analysis 4 – Multiple Imputation: Telehealth visits only

At a visit, if a subject has an in-clinic assessment and no telehealth assessment, the visit is assumed to be MAR.

- Intermittent missing data for telehealth visits will first be imputed using the MCMC method, appropriate for non-monotonic missing data.
- Data missing after subjects discontinue the study early will be imputed using a regression, appropriate for monotone missingness. At each time point, missing data will be assumed to follow a distribution similar to scores for subjects who are still in the study and randomized to the same treatment group.

The MI procedure will be employed as described for sensitivity analysis. MMRM will be applied to the multiply imputed data sets, as described for the key secondary estimand. The estimates from these models will then be combined using Ruben's (1987) rule and analysed to provide the endpoint estimates and summarized

Sensitivity analysis 5 – Completers

The MMRM will be applied to all subjects who have completed the study to Visit [redacted]

Sensitivity Analysis 6 – Rank Analysis

The CGI-I at EOT will be ranked with any tied observations being assigned an average rank. The sum of the ranks for each treatment will be calculated. The non-parametric Wilcoxon rank-sum test will be utilised at a 2-sided 5% level of significance to test the null hypothesis that the distribution of change from baseline for tideglusib and placebo are equal against the alternative hypothesis that the distribution of change from baseline for tideglusib and placebo are not equal. The Hodges-Lehmann estimate of the difference between the two treatments and 95% confidence interval will also be calculated.

Sensitivity analysis 7 – Multiple Imputation: MNAR control-based pattern

For this analysis, in-clinic and telehealth data will be used as for the primary analysis.

- Intermittent missing data for in-clinic visits will first be imputed using the FCS method, appropriate for non-monotonic missing data.

- Data missing after subjects discontinue the study early will be imputed using a regression, appropriate for monotone missingness. At each time point, missing data will be assumed to follow a distribution similar to scores for subjects who are still in the study and randomized to placebo.

The MI procedure will be employed as described for sensitivity analysis 1. The MMRM will be applied to the multiply imputed data sets, as described above for the primary estimand. The estimates from these models will then be combined using Ruben's (1987) rule and analysed to provide the endpoint estimates and summarized.

Sensitivity analysis 8 – Multiple Imputation: MNAR delta-adjustment pattern

This analysis will be done when there is a numerically positive treatment effect. For this analysis, in-clinic and telehealth data will be used as for the primary analysis.

- Intermittent missing data for in-clinic visits will first be imputed using the FCS method, appropriate for non-monotonic missing data.
- Data missing after subjects discontinue the study early will be imputed using a regression, appropriate for monotone missingness. At each time point, missing data will be assumed to follow a distribution similar to scores for subjects who are still in the study and randomized to the same treatment group. However, a fixed 5% worsening (delta-adjustment) to the outcome in the tideglusib group will then be applied.
- In the event of a statistically significant treatment effect, several delta-adjustments will be performed until non-significance occurs. This will be used to identify how much of a worse response would need to have occurred in the subjects with missing data in order to have overturned the significant result (a tipping-point analysis).

The MI procedure will be employed as described for sensitivity analysis. MMRM will be applied to the multiply imputed data sets, as described above for the primary estimand. The estimates from these models will then be combined using Ruben's (1987) rule and analysed to provide the endpoint estimates and summarized.

<u>Supplementary Analysis</u>

- The key secondary efficacy analysis will be repeated on the PPS and FAS.
- The proportion of subjects who improve, with CGI-I scores of 1-3 (Very much improved to minimally improved) vs. 4-7 (no change to very much worse) will be summarised and compared using a Fisher's Exact test.
- The proportion of subjects who worsen, with CGI-I scores of 5-7 (minimally worse to very much worse) vs. 1-4 (Very much improved to no change) will be summarised and compared using a Fisher's Exact test.
- The proportion of subjects with robust improvement, with CGI-I scores of 1-2 (Very much improved and much improved) vs. 3-7 (minimally improved to very much worse) will be summarised and compared using a Fisher's Exact test.
- Subgroup analyses will be performed by key baseline characteristics. The balance between treatment groups will be assessed and if any are evaluated as not being consistent a respective covariate will be included in the MMRM. For each subgroup, the main effect, treatment-by-subgroup interaction terms

and treatment by visit by subgroup interaction will be added to the model. Any interactions that are statistically significant at the 15% level for the end of treatment visit will have their nature described. These models will be used to estimate treatment comparisons within the subgroups that correspond with the sub-grouping factor. The results from the analysis are to be presented as described above in tables and/or figures. Subgroups of interest include but are not limited to:

- Race (White, Non-White)
- Sex (Male, Female)
- Age group ([redacted])
- Improvement in clinician-completed in-clinic CDM1-RS during the [redacted] run-in period (Yes, No).

Summaries and Figures

Summary statistics of observed values will be presented as both continuous and categorical by visit and treatment for the in-clinic and telehealth visits separately.

By subject line plots of the observed CGI-I results will be presented over time by treatment group. In addition, line plots will present the adjusted LS means and associated SEs for the observed values over time, as estimated by the MMRM model, by treatment group.

For the 3 occasions in the study (Visit [redacted] and [redacted] where dual telehealth and in-clinic assessments are conducted, summary statistics of the difference between the methods (in-clinic – telehealth) will be calculated and summarized using descriptive statistics over time. The intra-rater correlation will also be presented for each visit. A scatterplot for each visit will compare the observed value for the two methods for each subject, with in-clinic assessments being presented on the y-axis and telehealth on the x-axis. A linear regression line separately for each treatment will be included.

A line plot will present the observed values by visit, with in-clinic and telehealth assessments being presented on the same plot. Bar charts of the CGI-I frequency over time will be produced for the in-clinic and telehealth interviews separately. The CGI-I observed value (1 to 7) for Visits [redacted] and [redacted] will be presented on the x-axis and the number of subjects will be presented on the y-axis. The proportion of subjects achieving a CGI-I of 1, 2 or 3 will be presented similarly.

Data for in-clinic and telehealth assessments will be listed separately, with whether a telehealth assessment is within the pre-defined visit window for imputation as an in-clinic assessment being annotated.

6.13.3 Additional Secondary Efficacy Endpoints

6.13.3.1 Top 3 Caregiver Concerns Visual Analogue Scale (VAS)

Caregivers will be asked to rate three causes for concern by drawing a vertical mark on a 10 cm long VAS with anchors of “not at all severe” at the left end and “very severe” at the right end. The 3 concerns related to the subject’s myotonic dystrophy will be chosen and rated at Visit [redacted] and should not change for the duration of the study. The Top 3 Caregiver Concerns VAS is assessed at Visits [redacted] and [redacted] (Follow-up) and may be completed at home if required due to a COVID-19 related reason.

The Top 3 Caregiver Concerns VAS total score will be derived as the total of the VAS scores for the three concerns at each Visit regardless of any changes to the concerns rated. If fewer than three scores are available, the total score will be set to missing.

A sensitivity analysis will be conducted where the only the same concerns as recorded at Visit [redacted] may be used in the derivation of the total score. If not all of the three concerns are rated at a particular visit, the total score will be set to missing.

A MMRM model will be fitted to the change from baseline to Visit [redacted] and [redacted] in the the Top 3 Caregiver Concerns VAS total score. The model will be fitted and results presented as described for the primary estimand in section 6.13.1.1.

The observed Top 3 Caregiver Concerns VAS total score and corresponding change from baseline will be summarised over time. In addition, the VAS total score at Visits [redacted] and [redacted] and the change from Visit [redacted] to [redacted] will be summarised.

The observed Top 3 Caregiver Concerns VAS total score will also be presented over time using by subject line plots by treatment group. In addition, line plots will present the adjusted LS mean changes from baseline and SEs over time, as estimated by the MMRM model, by treatment group.

The Top 3 Caregiver Concerns data will be listed. The listing will include change from baseline for the individual concerns VAS scores as well as the total VAS scores.

6.13.3.2 Caregiver-Completed Congenital DM1 Rating Scale (CC-CDM1-RS)

The CC-CDM1-RS is completed by the caregiver to score the symptom severity of the domains that are clinically relevant in Congenital DM1. This scale is analogous to the Clinician Completed CDM1-RS. The severity of the caregiver's concern in each domain is scored by using a 5-point Likert Scale. Scores range from 0 = Not present to 4 = Very severe. The CC-CDM1-RS is assessed at Visits [redacted] and [redacted] and may be completed at home if required due to a COVID-19 related reason.

The CC-CDM1-RS total score will be derived at each visit as the total score based on 11 domain ratings, with a higher score indicating worse symptom severity. If one or more of the 11 items is missing then the total score will be set to missing.

An MMRM model will be fitted to the change from baseline to end of treatment in the CC-CDM1-RS in a similar manner to the primary estimand (see Section 6.13.1). The MMRM model will contain treatment, age group, visit and treatment by visit interaction as fixed effects, baseline CC-CDM1-RS as a covariate.

The CC-CDM1-RS total score and corresponding change from baseline will be summarised over time. In addition, the CC-CDM1-RS total score at Visits [redacted] and [redacted] and the change from Visit [redacted] to [redacted] will be summarised.

The change from baseline in CC-CDM1-RS total score will be presented over time using by subject line plots by treatment group. In addition, line plots will present the adjusted LS mean changes from baseline and SEs over time, as estimated by the MMRM model, by treatment group.

6.13.3.3 Clinical Global Impression – Severity Scale (CGI-S)

The CGI-S is a 7-point Likert type scale that requires the clinician to rate the severity of the subject's illness at the time of assessment, relative to the clinician's past experience with subjects who have the same diagnosis. Considering total clinical

experience, a subject is assessed on CGI-S at the time of rating from 1 = normal, not at all ill to 7= extremely ill.

The CGI-S is administered at Visit [redacted] and [redacted]. A negative mean change from baseline indicates an improvement since baseline, whilst a positive change from baseline indicates a worsening. The CGI-S may be assessed via telehealth instead of an in-clinic interview if required due to a COVID-19 related reason. If a subject's in-clinic visit is unavailable, the data will be imputed with the corresponding telehealth assessment.

An MMRM model will be fitted to the change from baseline to end of treatment in the CGI-S in a similar manner to the primary estimand (see Section 6.13.1). The MMRM model will contain treatment, age group, visit and treatment by visit interaction as fixed effects, baseline CGI-S and age group as covariates.

Summary statistics of observed values will be presented as both continuous and categorical with the associated change from baseline and categorical by visit and treatment for the in-clinic and telehealth visits separately. For subjects who have both an in-clinic and telehealth assessment, change from baseline in CGI-S will be calculated for both methods and summarised separately.

The CGI-S will be presented over time using by subject line plots by treatment group. In addition, line plots will present the adjusted LS mean changes from baseline and SEs over time, as estimated by the MMRM model, by treatment group.

For the 3 occasions in the study (Visit [redacted]) where dual telehealth and in-clinic assessments are conducted, summary statistics of the difference between the methods (in-clinic – telehealth) will be calculated and summarized using descriptive statistics over time. The intra-rater correlation will also be presented for each visit. A scatterplot for each visit will compare the observed value for the two methods for each subject, with in-clinic assessments being presented on the y-axis and telehealth on the x-axis. A linear regression line separately for each treatment will be included. A line plot will present the mean at each visit, with in-clinic and telehealth assessments being presented on the same plot. Bar charts of the CGI-S frequency over time will be produced for the in-clinic and telehealth interviews separately. The CGI-S score (1 to 7) for Visits [redacted] and [redacted] will be presented on the x-axis and the number of subjects will be presented on the y-axis.

Data for in-clinic and telehealth assessments will be listed separately.

6.13.3.4 Independent Central Rater of CDM1-RS and CGI

The in-clinic and telehealth CDM1-RS and CGI interviews at Visits [redacted] and [redacted] are video recorded. In addition to the investigator ratings of the CDM1-RS and CGI, an independent central rater will also rate the CDM1-RS, CGI (CGI-I and CGI-S) for both the in-clinic and telehealth interviews using the video recordings.

These measures will be summarised and analysed in a similar way to the corresponding clinician completed versions.

For each visit summary statistics of the difference between the methods (independent rater in-clinic – clinician rated in-clinic) will be calculated and summarized using descriptive statistics over time. The intra-rater correlation will also be presented for each visit. A scatterplot for each visit will compare the observed value for each method for each subject, with independent rater in-clinic assessments being presented on the y-axis and clinician rated in-clinic on the x-axis. A linear regression line will be included.

This will be repeated with the in-clinic measurements replaced by the telehealth measurements.

The Independent Central Rater data for the CDM1-RS and CGI will be listed.

6.13.3.5 10 Metre Walk/Run Test

A 10-metre ambulatory test will be performed at the subject's preferred walking speed and then at the fastest speed possible. For each speed, the test should be repeated until 3 valid measures are obtained. The mean of repeated measurements for each speed will be derived for analysis. In the event that a subject does not complete a walk/run test, then the derivation will be based on the observed mean of the non-missing repeat assessments. The 10-metre walk/run test is assessed at Visits [redacted] and [redacted]

A MMRM will be fitted in a similar manner to the primary estimand (see Section 6.13.1) to the change from baseline over time in the time taken (seconds) to complete the 10-metre walk/run test for the mean preferred speed and fastest speed. Each model will include the baseline value and age group as covariates, treatment group and visit as fixed effects, and the treatment-by-visit interaction.

For preferred speed and the fastest speed separately, the mean (of the repeat assessments at a visit) in the time taken (seconds) to complete the 10-metre walk/run test will be presented using by subject line plots over time by treatment group. In addition, line plots will present the adjusted LS mean changes from baseline and SEs over time, as estimated by the MMRM model, by treatment group for the time taken (seconds) to complete the 10-metre walk/run test.

The 10-metre walk/run test data will be listed.

6.13.4 Exploratory Endpoints

6.13.4.1 Dual-energy X-ray Absorptiometry (DXA) Whole Body Scan of Lean Muscle Mass

The DXA scan measures bone mineral density, lean muscle mass, fat content, and total body composition. The primary assessment of interest in this study, from an efficacy perspective, is lean muscle mass. Subjects will have their lean muscle mass determined by a whole body DXA scan and the following parameters recorded in grams; arms, legs and total. The DXA scan is performed at Visits [redacted] and [redacted]

An ANCOVA will be fitted to the change from baseline to end of treatment in the DXA scan total lean muscle mass (g). The ANCOVA will include the baseline value as a covariate and treatment group and age group as fixed effects. Adjusted LS Means for the treatment group means at the end of treatment estimated by the above model will be presented, together with the associated SE, 95% CIs and 2-sided p-values.

DXA scan total lean muscle mass (g) and change from baseline will be summarised over time using summary statistics.

DXA scan data will be listed.

6.13.4.2 Measurement of lip strength (via lip force meter)

Lip strength will be measured using a mouthguard adaptation for the digital force meter. Subjects will insert the mouthguard between their incisors and their lips and hold as force is applied for 10 seconds with increasing force until the mouthguard is

dropped. Measurements will be repeated until 3 valid times are obtained. Lip strength is assessed at Visits [redacted] and [redacted]

The mean of repeated measurements will be derived for analysis. In the event that a subject does not complete a lip strength test, then the derivation will be based on the observed mean of the non-missing repeat assessments.

A MMRM will be fitted to the change from baseline over time in the mean lip strength (Newtons) in a similar manner to the primary estimand (see Section 6.13.1). The model will include the baseline value as a covariate, treatment group, age group and visit as fixed effects, and the treatment-by-visit interaction.

The mean (of the repeat assessments at a visit) in lip strength will be presented using by subject line plots over time by treatment group. In addition, line plots will present the adjusted LS mean changes from baseline and SEs over time, as estimated by the MMRM model, by treatment group.

The lip strength data will be listed.

6.13.4.3 Congenital and Childhood Myotonic Dystrophy Health Index (CC-MDH1) Parent Proxy Instrument

The CC-MDHI Parent Proxy is a 19 category, 128-item Likert scale that covers multiple aspects of the specific phenotype of Congenital DM1, intended for completion by the caregiver of the subject. Each item offers 6 levels of severity ranging from “he/she doesn’t experience this” to “it affects his/her life severely”. The CC-MDHI will be administered at Visits [redacted] and [redacted]. Each item will be scored 0 (corresponding to “he/she doesn’t experience this” to 5 (corresponding to “it affects his/her life severely”).

The CC-MDHI total score will be derived at each visit as the total score based on the 128 items, and the score within each of the 19 categories will also be calculated. A higher score indicates worse symptoms or severity. If one or more items is missing, the scores will be imputed based on the values of the other items in that category in which the item is missing by taking the average of the non-missing items and multiplying by total number of items (non-missing and missing) included in the score. This will be done provided at least half of the items within a particular category are non-missing.

An ANCOVA will be fitted to the change from baseline to end of treatment in the CC-MDHI total score. The same will be done for the total scores in each of the 19 categories. The model will be fitted and results presented as described in section 6.13.4.1 for the DXA scan.

Summary statistics of observed values with corresponding change from baseline will be presented as continuous by visit and treatment. Line plots will present the adjusted LS mean changes from baseline to end of treatment and SEs, as estimated by the ANCOVA, by treatment group.

The CC-MDHI data will be listed.

6.13.4.4 Autism Behavior Inventory - Clinician (ABI-C)

The ABI-C is a 14-item rating scale to assess the core features of Autism Spectrum Disorder (ASD) as well as common associated behaviors. When completing the ABI-C, the clinician is asked to rate the overall severity/level of impairment of each item over the past week on a 7-point Likert type scale. In general, the ratings correspond

to the following: 1. None, No symptoms present to 7. Very Severe, Persistent interference with function or adaptation.

The clinician will rate the subject based on both behavioral observation and clinician interview. The ABI-C will be administered at Visits [redacted] and [redacted] and may be assessed via telehealth if required due to a COVID-19 related reason.

The ABI-C total score will be derived at each visit as the total score based on the 14 items, with a higher score indicating worse ASD severity. If one or more of the 14 items is missing then the total score will be set to missing. If a subject's in-clinic visit is unavailable, the data will be imputed with the corresponding telehealth assessment.

The ABI-C core autism symptoms subscore will be derived at each visit as the total score based on the first 8 items, with a higher score indicating worse ASD severity. If one or more of the 8 items is missing then the total score will be set to missing. If a subject's in-clinic visit is unavailable, the data will be imputed with the corresponding telehealth assessment if available.

A MMRM will be fitted to the change from baseline over time in the ABI-C total score and the ABI-C core autism symptoms subscore in a similar manner to the primary estimand (see Section 6.13.1.1). The model will include the baseline value as a covariate, treatment group, age group and visit as fixed effects, and the treatment-by-visit interaction.

Summary statistics of observed values with corresponding change from baseline will be presented as continuous by visit and treatment. Line plots will present the adjusted LS mean changes from baseline to end of treatment and SEs, as estimated by the MMRM model, by treatment group.

The ABI-C data will be listed.

6.13.4.5 Vineland Adaptive Behavior Scale

The Vineland Adaptive Behavior Scales – [redacted] (Vineland [redacted] measures personal and social skills needed for everyday living. The scales are organized into a [redacted]: [redacted]. A subject's overall level of adaptive functioning is described by the Adaptive Behavior Composite score. There are several items in each of the [redacted], and each item is scored as a [redacted] The scores are summed for each domain to generate a raw score for each domain. These raw scores are also converted into standard scores. The sum of the standard scores for the [redacted] is the Adaptive Behavior Composite score.

The Vineland[redacted] will be administered at Visit [redacted] and [redacted] and will be assessed via telehealth if required due to a COVID-19 related reason. If a subject's in-clinic visit is unavailable, the data will be imputed with the corresponding telehealth assessment.

For each raw domain score, standardised domain score and the Adaptive Behavior Composite score, an ANCOVA will be fitted to the change from baseline to end of treatment. The model will be fitted and results presented as described in section 6.13.4.1 for the DXA scan.

Summary statistics of observed values with corresponding change from baseline will be presented as continuous by visit and treatment. Line plots will present the adjusted LS mean changes from baseline to end of treatment and SEs, as estimated by the ANCOVA, by treatment group.

The Vineland Adaptive Behavior Scale data will be listed.

6.13.4.6 Quantitative myometric measure of hand grip strength

Grip strength is recorded bilaterally, first on the right hand followed by the left hand. For each hand 3 repeat assessments are recorded. For analysis purposes, the mean of the repeat assessments is taken. Hand grip strength will be tested at Visits [redacted] and [redacted]

For each hand, a MMRM will be fitted to the change from baseline over time in the mean hand grip strength (kg) in a similar manner to the primary estimand (see Section 6.13.1). Each model will include the baseline value a covariate, treatment group, age group and visit as fixed effects, and the treatment-by-visit interaction.

For each hand, the change from baseline will be summarised over time, with the observed mean at Visits [redacted] and [redacted] and the change from Visit [redacted] to [redacted] in the observed mean also summarised. The observed mean will also be presented using by subject line plots over time by treatment group separately for each hand. In addition, line plots will present the adjusted LS mean changes from baseline and SEs over time, as estimated by the MMRM model, by treatment group.

The grip strength data will be listed.

6.13.4.7 NIH Toolbox Cognition Battery

6.13.4.7.1 Dimensional Change Card Sort Test (DCCS)

The NIH Toolbox Cognition Battery: Dimensional Change Card Sort Test (DCCS) is a measure of cognitive flexibility and attention in individuals aged 3 years and older. Two target pictures are presented that vary along two dimensions (e.g., shape and color). Subjects are asked to match a series of bivalent test pictures (e.g., yellow balls and blue trucks) to the target pictures, first according to one dimension (e.g., color) and then, after a number of trials, according to the other dimension (e.g., shape). The relevant dimension for sorting is indicated by a cue word (e.g., “shape” or “color”) that appears on the screen for all subjects and that, for young children ages 3-11 years, is also spoken by a pre-recorded audio file. Each administration of the test produces a raw score, computed score, uncorrected standard score, an age-corrected standard score, national percentile and full corrected T-score. The DCCS will be administered at Visit [redacted] and [redacted].

Separately for the raw score, computed score and uncorrected standard score an ANCOVA will be fitted to the change from baseline to end of treatment. The model will be fitted and results presented as described in section 6.13.4.1 for the DXA scan.

Summary statistics of observed values with corresponding change from baseline will be presented as continuous by visit and treatment. Line plots will present the adjusted LS mean changes from baseline to end of treatment and SEs, as estimated by the ANCOVA, by treatment group.

The DCCS data will be listed.

6.13.4.7.2 Picture Sequence Memory Test (PSMT)

The NIH Toolbox Cognition Battery: Picture Sequence Memory Test (PSMT) is a measure developed for the assessment of episodic memory for ages 3-85 years. It involves recalling an increasingly lengthy series of illustrated objects and activities that are presented in a particular order on the iPad screen, with corresponding audio-

recorded phrases. The sequence of pictures appears one at a time in the center of the computer screen in a fixed order. As each picture appears, a recording briefly describes its content. This continues until all pictures in a sequence have been displayed and placed in their proper positions. Once all the pictures in the sequence are displayed, the pictures then are placed in a random spatial array at the center of the screen. The subjects are asked to recall the sequence of pictures demonstrated over two learning trials; sequence length varies from 6-18 pictures, depending on age. Subjects are given credit for each adjacent pair of pictures they correctly place. The test takes approximately seven minutes to administer. Each administration of the test produces a raw score, computed score, uncorrected standard score, an age-corrected standard score, national percentile and full corrected T-score. The DCCS will be administered at Visit [redacted] and [redacted].

Separately for the raw score, computed score and uncorrected standard score an ANCOVA will be fitted to the change from baseline to end of treatment. The model will be fitted and results presented as described in section 6.13.4.1 for the DXA scan.

Summary statistics of observed values with corresponding change from baseline will be presented as continuous by visit and treatment. Line plots will present the adjusted LS mean changes from baseline to end of treatment and SEs, as estimated by the ANCOVA, by treatment group.

The PSMT data will be listed.

6.13.4.8 Peabody Picture Vocabulary Test (PPVT)

The PPVT–4 scale is a norm-referenced instrument for measuring the receptive (hearing) vocabulary of children and adults. It contains training items and 228 test items, each consisting of four full-color pictures as response options on a page. For each item, the examiner says a word, and the examinee responds by selecting the picture that best illustrates that word's meaning. Each administration of the test produces a raw score as well as a standard score, derived from the number of responses attempted and the number of correct responses. The percent of correct responses is also reported. The PPVT will be performed at Visits [redacted] and [redacted].

An ANCOVA will be fitted to the change from baseline to Week [redacted] in the PPVT age-based standard score. The model will be fitted and results presented as described in section 6.13.4.1 for the DXA scan. The model will include the baseline value as a covariate and treatment group and age group as fixed effects.

The observed raw score, age-based standard score, percentile rank, age equivalent in years and months and corresponding change from baseline over time will be summarised. The above PPVT endpoints (apart from the percentile rank and age equivalent endpoints) will also be presented over time using by subject line plots by treatment group. In addition, line plots will present the adjusted LS mean changes from baseline to end of treatment and SEs, as estimated by the ANCOVA, by treatment group for the PPVT raw score and age-based standard score.

The PPVT data will be listed.

6.13.5 [redacted] levels

A blood sample will be taken for the [redacted] analysis will be taken at Visits [redacted] and [redacted]

An MMRM model will be fitted to the change from baseline to end of treatment in [redacted] in a similar manner to the primary estimand (see Section 6.13.1). The MMRM model will contain treatment, age group, visit and treatment by visit interaction as fixed effects, baseline [redacted] as a covariate.

[redacted] and corresponding change from baseline will be summarised over time.

The change from baseline in [redacted] will be presented over time using by subject line plots by treatment group. In addition, line plots will present the adjusted LS mean changes from baseline and SEs over time, as estimated by the MMRM model, by treatment group.

6.13.6 [redacted]

A 5ml blood sample will be taken for [redacted] analysis, and for future, potential [redacted] analysis at Visits [redacted] and [redacted]. The analysis of these data is outside the scope of this SAP.

6.13.7 [redacted]

At Visit [redacted] and Visit [redacted] an [redacted] may be performed. The analysis of these data is outside the scope of this SAP.

6.14 Multiplicity

For all analyses, the effects of the weight adjusted tideglusib 1000 mg dose versus placebo will be estimated at EOT and presented along with associated 95% two-sided confidence intervals and p-values. The comparisons may also be performed at other time points; however the p-values should be reviewed with caution due to potential multiplicity issue (high probability of chance-alone outcomes).

An overall false-positive rate of 5% level for the key secondary efficacy estimand will be maintained, in that no significance of key secondary estimand will be claimed unless the primary statistical analysis is significant at the 5% level.

All other secondary endpoints and the supportive analyses will be considered as supportive evidence of efficacy and will be analyzed without any procedures to account for multiple comparisons.

6.15 Concordant Trend Analysis

A supporting exploratory concordant trend analysis will be performed as described by Berry-Kravis et al., 2020; Glaze et al., 2017; Jessup et al., 2011. The approach is based on the concept that concordant trends in multiple efficacy domains may provide strong evidence of efficacy in small studies, especially when no established outcome measure exists, as for Congenital DM1. The details of this analysis are presented in the Appendix 1

The efficacy variables in this study were categorised into four different efficacy domains as shown in Table 2:

Table 2: AMO-02-MD-2-003 Outcome Variables by Efficacy Domains

[redacted]

[redacted]

The concordant trend analysis is based on a combination of group-level and subject-level analyses of the above variables. The study outcome may be indicative of overall biologic activity/efficacy if efficacy is detected for group- and/or subject-level analyses with at least a corresponding trend in the other (group- or subject-level) analyses. A corresponding trend is defined as numerical superiority of tideglusib to placebo as described for the group- and subject-level analyses but regardless of the p-values obtained for a given comparison.

The success definition is based on a combination of criteria to control the probability of a false-positive study outcome. Appendix 1 provides further details of the success criteria and analysis approach.

6.15.1 Permutation Test

If study outcomes meet or exceed the success definition, a permutation test will be performed to determine the probability of obtaining success by chance alone. An attractive aspect of the permutation test is that it preserves the correlation structure in the study data, takes multiplicity into account and does not require additional assumptions.

If the concordant trend analysis meets the pre-specified requirements for overall efficacy specified in Appendix 1, a permutation test will be performed. The permutation test will be conducted under the assumption that there is no difference between tideglusib and placebo, thereby determining the false-positive rate on the actual study results. Randomly simulated allocations of subjects to tideglusib and placebo will be repeated 1000 times and positive outcomes will be counted. The false-positive rate for overall efficacy will be determined by:

$$False\ positive\ rate = \left(\frac{Total\ positive\ outcomes\ in\ 1000\ simulations}{1000}\right)$$

If the actual study outcomes exceeded the minimal requirements for defining overall biologic activity/efficacy as specified, then the permutation test will be performed to determine the false-positive rate based on these actual study results and not based on minimal requirements.

6.16 Pharmacokinetics

Pharmacokinetic blood samples will be collected for measurement of tideglusib and its primary metabolite, NP04113. PK samples will be collected at Visits [redacted] and [redacted] (corresponding to Weeks [redacted] and [redacted]), with [redacted] will be collected at Visit [redacted] and [redacted] samples at Visit [redacted] The time of food and type of food before and after dosing is recorded for Visits where PK is collected.

Plasma concentrations of tideglusib and its primary metabolite NP04113 will be listed and summarised over time using the PK analysis set. The descriptive statistics will include the geometric mean and the number of subjects with a BLQ concentration. Concentrations reported as below the limit of quantification (BLQ) will be taken as zero for linear plots and descriptive statistics, and equal to the lower limit of quantification (LLOQ) for semi-logarithmic plots and the calculation of geometric means.

Individual plasma concentration profiles will be plotted both on the original scale and on the log scale separately. Mean (±SD) and median plasma concentration profiles will be plotted on the original scale and on the log scale.

A population pharmacokinetic analysis will be conducted. This is outside the scope of this SAP.

6.17 Safety Evaluation

The safety and tolerability of tideglusib will be assessed based on adverse events, and safety evaluations including clinical laboratory evaluations, vital signs and ECGs. This will be done for the SAF set. In addition, [redacted] run-in adverse events will be summarized for the RIS set.

6.17.1 Adverse Events

Adverse events (AEs) will be coded using the MedDRA dictionary version applicable at eCRF go-live. The version used will be indicated in the data summaries and listings.

A Treatment-Emergent AE (TEAE) is defined as an AE that started on or after the start of the first dose of double-blind, randomized study treatment. If adverse event dates are incomplete and it is not clear whether the adverse event was treatment-emergent, it will be assumed to be treatment-emergent.

A Non-Treatment Emergent Adverse Event (NTEAE) is defined as an AE that started before the administration of randomized study treatment.

A “double-blind treatment period” TEAE is defined as an AE that started on or after the start of randomized study treatment but before 2 days after randomized treatment was discontinued. A “post treatment period” TEAE is defined as an adverse event that started from 2 days after the randomized study treatment being discontinued.

A treatment related TEAE is defined as a TEAE that is related to the randomized study treatment. If the TEAE has a missing relationship it is assumed to be related to the randomized study treatment for analysis purposes.

The following summaries of the total number of “double-blind treatment period” TEAEs, “post treatment period” TEAEs and all TEAEs (both “double-blind treatment period” and “post treatment period” TEAEs) will be provided at the event and subject levels as indicated:

- TEAEs (number of events and number of subjects).
- Serious TEAEs (number of events and number of subjects).
- Serious randomized study treatment-related TEAEs (number of events and number of subjects).
- TEAEs by severity (mild/moderate/severe) (number of events and number of subjects).
- TEAEs by relationship to randomized study treatment category (unrelated/related) (number of events and number of subjects).
- TEAEs leading to discontinuation from the study (number of subjects only).
- Randomized study treatment-related TEAEs leading to discontinuation from the study (number of subjects only).
- TEAEs leading to death (number of subjects only).

The following summaries of NTEAEs will be provided at the event and subject levels as indicated:

- NTEAEs (events and subjects).
- NTEAEs by severity (mild/moderate/severe) (events and subjects).
- Serious NTEAEs (events and subjects).
- NTEAEs leading to withdrawal (subjects only).
- NTEAEs leading to death (subjects only).

In the above summaries, if a subject experienced more than one NTEAE/TEAE, the subject will be counted once using the most related event for the “by relationship to treatment” summary and the worst severity for the “by severity” summary. The Enrolled analysis set will be used to summarize NTEAEs.

A table of TEAEs by System Organ Class (SOC) and Preferred Term (PT) will be presented for the following:

- Events and subjects for “double-blind treatment period” TEAEs, “post treatment period” and all TEAEs.
- Events and subjects for “double-blind treatment period” treatment-related TEAEs, “post treatment period” and all treatment-related TEAEs.
- Events and subjects for “double-blind treatment period” serious TEAEs, “post treatment period” and all serious TEAEs.
- Events and subjects for all NTEAEs.

For all of the above, SOC and PT will be presented in decreasing frequency of the total number of subjects with TEAEs. If a subject experienced more than one TEAE, the subject will be counted once for each SOC and once for each PT.

A table of TEAEs by severity, System Organ Class (SOC) and Preferred Term (PT) will also be presented. If a subject experienced more than one TEAE, the subject will be counted once for each SOC and once for each PT at the most severe event.

Adverse event data will be listed in full and this will also include a treatment emergent flag, a double-blind treatment period TEAE flag, a post treatment period TEAE flag, the time of onset and cessation of event relative to first dosing of randomized study treatment and duration of AE.

TEAEs leading to death or AEs leading to death which will also include a treatment emergent flag, serious TEAEs and TEAEs leading to discontinuation of study drug will also be listed.

6.17.2 Clinical Laboratory Evaluation

Observed values and change from baseline in [redacted], [redacted] and [redacted] assessments will be summarized over time. At baseline the change from run-in baseline will also be presented. If the test results are reported in categorical format, the results will be summarized by subject counts and percentage for each category.

Each parameter will be classed as low, normal, high based on the reference ranges. Shift tables in relation to the normal range from baseline over time will be presented.

[redacted] ([redacted]), virology and urinalysis data will be listed separately including change from baseline, reference ranges flagging all out of range values and their clinical significance. Separate listings of out of range laboratory measurements recorded throughout the study will be provided. Local laboratory results will be presented on the same listings, where these are collected.

Spaghetti plots of [redacted] and [redacted] will be produced containing a separate line for each subject, with the measured value appearing on the y-axis and time (days since first administration of randomized study treatment) on the x-axis. The relevant reference range and 1 x, 2x and 3x Upper Limit of Normal (ULN) will be overlaid on the plot as a reference line.

6.17.3 Vital Signs

Observed values and changes from baseline for weight, pulse rate, respiration rate, systolic blood pressure, diastolic blood pressure and body temperature will be summarized over time. At baseline the change from run-in baseline will also be presented.

Parameters will be presented in the same order as the eCRF.

All vital sign data will be listed including reference ranges flagging all out of range values and changes from baseline/run-in baseline. Changes from baseline in vital signs of clinical concern recorded throughout the study will also be flagged. Changes of clinical concern are defined as a change from baseline in systolic blood pressure of more than 40 mmHg, a change from baseline in diastolic blood pressure of more than 20 mmHg, or a change from baseline in heart rate of more than 30 bpm.

A listing of subjects with at least one vital sign of clinical concern recorded throughout the study will be provided.

6.17.4 Electrocardiography

All ECGs will be performed using equipment provided by the central ECG provider and all ECGs performed will be transmitted electronically to the central ECG provider for interpretation. For any visits completed in a subject's home due to a COVID-19 related reason, the standard 12-lead ECG will be replaced with a 6-lead ECG using equipment provided by the central ECG vendor. The investigator will perform the initial interpretation of the ECG immediately after collection to ensure the safety of each subject. The central ECG provider will also provide an interpretation.

ECG parameters will include heart rate, PR interval, QRS interval, QT interval, and QTcF and QTcB intervals (QT interval corrections will be those values provided by the central ECG provider).

ECG observed values and changes from baseline/run-in baseline will be summarized over time by timepoint (if applicable), Type (12-lead or 6-lead) for each parameter.

For overall interpretations if more than one value is recorded per assessment, then the most severe of the respective readings will be taken.

The number and percentage of subjects meeting the following categories will be tabulated over time by timepoint (if applicable):

Table 3: AMO-02-MD-2-003 ECG Criteria

Parameter	Criteria
QT and QTcF/QTcB intervals	<450 msec ≥ 450 msec and < 480 msec * ≥ 480 msec and < 500 msec * ≥ 480 msec * ≥ 500 msec *
Increase from Baseline/Run-in baseline in QT and QTcF/QTcB intervals	< 30 msec ≥ 30 msec and < 60 msec * ≥ 60 msec *

* values of potential clinical importance

In addition, the overall interpretation according to the investigator of the ECG (Normal, Abnormal Not Clinically Significant, and Abnormal Clinically Significant) will be summarized over time.

Shift in ECG overall interpretation according to the investigator over time for normal, abnormal not clinically significant, abnormal clinically significant and missing results will be summarised.

Similar summaries for the central ECG provider interpretation will also be made.

Parameters will be presented in the same order as the CRF.

All ECG results will be listed including changes from baseline/run-in baseline values. QT, QTcF and QTcB values of ≥ 480 msec and ≥ 500 msec and increases from baseline ≥ 60 msec will be flagged.

6.17.5 Physical Exam

Details of timings of physical examinations and the presence or absence of clinically significant findings will be listed.

6.17.6 Pregnancy Test

Pregnancy test details will be listed.

6.18 Changes from the Protocol Planned Analysis

- The Enrolled and Run-in analysis sets are not mentioned in the protocol and have been added to this SAP.
- The protocol defined ITT analysis set derivation has been updated in this SAP following feedback from the FDA.
- The evaluation of the measurement properties of the CDM1-RS are not included in the protocol and have been added to this SAP.

7 REFERENCES

Berry-Kravis E, Horrigan JP, Tartaglia N, Hagerman R, Kolevzon A, Erickson CA, Hatti S, Snape M, Yaroshinsky A, Stoms G, FXS-001 Investigators, Glass L, Jones NE. (2020). A Double-Blind, Randomized, Placebo-Controlled Clinical Study of Trofinetide in the Treatment of Fragile X Syndrome. Pediatric Neurology. 110: 30-41

Glaze DG, Neul JL, Percy A, Feyma T, Beisang A, Yaroshinsky A, Stoms G, Zuchero D, Horrigan JP, Glass L, Jones NE. (2017) A Double-Blind, Randomized, Placebo-Controlled Clinical Study of Trofinetide in the Treatment of Rett Syndrome. Pediatric Neurology. 76: 37-46.

Jessup M, Greenberg B, Mancini D, Cappola T, Pauly DF, Jaski B, Yaroshinsky A, Zsebo KM, Dittrich H, Hajjar RJ, on behalf of the Calcium Upregulation by Percutaneous Administration of Gene Therapy in Cardiac Disease (CUPID) Investigators. (2011) Calcium Upregulation by Percutaneous Administration of Gene Therapy in Cardiac Disease (CUPID) A Phase 2 Trial of Intracoronary Gene Therapy of Sarcoplasmic Reticulum Ca2+ -ATPase in Patients with Advanced Heart Failure. Circulation. 124:304-313.

Rubin, D.B. (1987), Multiple Imputation for Nonresponse in Surveys, New York: John Wiley & Sons, Inc.

Koch, G.G., Tangen, C.M., Jung, J.-W. and Amara, I.A. (1998). Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. Stat Med., 17(15-16): 1863-92.

U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER) (February 2019). Guidance for Industry: Rare Diseases: Common Issues in Drug Development Guidance for Industry. FDA-2015-D-2818

U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER) (November 2019). Guidance for Industry: Adaptive Designs for Clinical Trials of Drugs and Biologics. FDA-2018-D-3124

U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER) (May 2021). Guidance for Industry: E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials. FDA-2017-D-6113

U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER) (May 2021). Draft Guidance for Industry: Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products. FDA-2019-D-0934

Zink, Richard & Koch, Gary. (2012). NParCov3: A SAS/IML Macro for Nonparametric Randomization-Based Analysis of Covariance. Journal of Statistical Software, 50:3 (1,17).

8 [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

9 [redacted]

[redacted]

A Randomized, Double-Blind Study to Evaluate the Efficacy and Safety of Tideglusib Versus Placebo for the Treatment of Children and Adolescents with Congenital Myotonic Dystrophy (REACH CDM)

Main Statistical Analysis Plan Addendum

Version: Final 2.0

Date: 22 August 2024

For [redacted] – Lead Statistician

Signed by:
[redacted]

For AMO Pharma Ltd

Signed by:
[redacted]

Signed by:
[redacted]

Contents

1 INTRODUCTION

This is an addendum to the final SAP V2.0 dated 22 February 2023, documenting additional *post-hoc* analyses. Full details of the analyses are detailed below.

2 CHANGES TO THE FINAL SAP

The following analyses will be performed for the ITT set unless otherwise stated.

2.1 Linear Trend Analysis

The following endpoints will be analysed using a random intercepts and slopes model:

- 10 Metre Walk/Run – Preferred Speed
- Hand Grip Strength – Dominant Hand
- Creatine Phosphokinase

The model fitted will include a term for a fixed effect of treatment by slope interaction and random effects of intercept and slope.

The model will utilize all non-missing baseline and on-treatment data. For post-baseline measurements the exact day of the measurement (rather than nominal week) will be used as the time element of the model. For baseline measurements the time will be set to [redacted]. For other measurements the time will be set to the [redacted] minus [redacted] since [redacted] begins at [redacted]

The model will be used to estimate the average slope (rate of change in the endpoint) and the mean value at [redacted] for each treatment group. The difference between the slopes and the means will be presented along with 95% confidence intervals for the differences and corresponding p-values.

The predicted mean over time for each treatment group based on the model will also be plotted.

The following SAS code, where y is the absolute value of the endpoint, usubjid is the subject number, trt01pn is the randomized treatment group and time is the actual time in days, will be used to fit the model:

[redacted]

This will also be done for the per protocol analysis set.

A table shell for this analysis is given in Section 2.7, Table 1.

Further analyses based on adding a quadratic term in time to the model may also be performed to assess if there is a significant departure from linearity. The possibility of log-transformation for creatine phosphokinase will also be explored.

2.2 Scorecard Analysis

A scorecard analysis will be performed by creating an efficacy score for each subject according to thresholds of clinically significant within-subject changes at Visit [redacted] as indicated in the literature. The scorecard analysis may also be known as the multi-domain responder index (MDRI).

[redacted]

For each endpoint, subjects will be scored [redacted] if their change exceeds the threshold in a beneficial direction, [redacted] if their change exceeds the threshold in a detrimental direction and [redacted] otherwise (including when the change cannot be calculated). An efficacy score, which is the total of these values, will then be calculated for each subject.

The efficacy score will then be compared between treatments using a t-test. Within treatment group Wilcoxon signed rank tests of the score will also be performed. Additionally, a listing of the subjects and each score and the total efficacy score will be produced.

The presentation of these analyses will be based on Version 1.0 of the Table, Figure and Listing Shells, Table 14.2.19.2 and Listing 16.2.6.18.

2.3 MMRM Analysis

An MMRM analysis will be performed for creatine phosphokinase and [redacted] absolute and ratio to baseline values as described for the primary endpoint in Version 2.0 of the Main Statistical Analysis Plan (Section 6.13.1.1). Shells for this analyses will be based on Version 1.0 of the Table, Figure and Listing Shells, Table 14.2.1.2.1.

The possibility of log-transformation for creatine phosphokinase will be explored. If this is required, results will be back-transformed and presented in terms of ratios to baseline within treatment group and ratio of geometric means between treatment groups.

This will also be done for the per protocol analysis set.

2.4 Responder Analysis

A responder analysis will be performed for each of the following endpoints. Proportions of responders will be compared between treatment groups using a Fisher's exact test.

- 10 Metre Walk/Run – Preferred Speed
- Hand Grip Strength – Dominant Hand
- Creatine Phosphokinase

- Peabody Picture Vocabulary Test – Raw Score
- Vineland Adaptive Behaviour Scale – [redacted] Raw Score
- Vineland Adaptive Behaviour Scale – [redacted] - Raw Score
- Vineland Adaptive Behaviour Scale – [redacted] - Raw Score
- Lip Strength
- DXA Scan – Total Lean Muscle Mass
- DXA Scan – Bone Mineral Density Total Z-Score
- NIH Toolbox Cognition Battery: Dimensional Change Card Sort Test (DCCS) – Computed Score
- NIH Toolbox Cognition Battery: Picture Sequence Memory Test (PSMT) – Computed Score
- [redacted]

For each endpoint, response is defined as improvement from baseline at Visit [redacted] ([redacted] and will be tested using the following thresholds: [redacted], [redacted] and [redacted]. In addition, clinically relevant thresholds as defined in section 2.2 will also be explored for 10 metre walk [redacted]), Peabody Picture Vocabulary Test – Raw Score ([redacted]), Hand Grip Strength – Dominant Hand ([redacted]), Vineland Adaptive Behaviour Scale – [redacted] – Raw Score ([redacted]). For the 10 metre walk, [redacted] and creatine phosphokinase a reduction is considered an improvement, while for the other endpoints an increase is considered an improvement. Subjects without a [redacted] and a Visit [redacted] ([redacted] value will be excluded from the analysis.

A test of the global null hypothesis of no treatment effect on any of these responder endpoints versus the one-sided alternative of an effect in favour of tideglusib on at least one of them will be performed using the test described by Rom (1992).

The presentation of these analyses will be based on Version 1.0 of the Table, Figure and Listing Shells, Table 14.2.1.2.17. A table shell for Rom's test is given in Section 2.7, Table 2.

2.5 Effect of Run-In Period Response

Subgroup analyses using MMRM will be performed for the following endpoints:

- Top 3 Caregiver Concerns VAS Total Score
- Caregiver-Completed Congenital DM1 Rating Scale (CC-CDM1-RS) Total Score

The subgroup will be response defined in the following two ways:

- A [redacted] improvement in the clinician-completed CDM1-RS total score (Yes/No) between Visit [redacted] and Visit [redacted]
- [redacted] (Yes/No)

The analysis will be performed as described in Version 2.0 of the Main Statistical Analysis Plan (Section 6.13.1.1). The presentation of these analyses will be based on Version 1.0 of the Table, Figure and Listing Shells, Table 14.2.1.2.16.

2.6 Correlation Analysis

The Pearson's and Spearman's rank correlation coefficient with the corresponding p-value will be presented for:

- In-clinic clinician CDM1RS and caregiver CDM1RS.
- 10 Meter Walk time (Preferred Speed) and CDM1RS Item #1 (Limitations on mobility or walking) at baseline and End of Treatment.
- Peabody Picture Vocabulary Test – Raw Score and CDM1RS Item #8 (Difficulty thinking) at baseline and End of Treatment.
- Handgrip strength (Dominant hand) and CDM1RS Item #2 (Problems with hands or arms) at baseline and End of Treatment.

A scatterplot for each visit will compare the observed value for the two methods for each subject. A linear regression line separately for each treatment will be included.

The presentation of these figures will be based on Version 1.0 of the Table, Figure and Listing Shells, Figure 14.2.6.3. A table shell for this analysis is given in Section 2.7, Table 3.

2.7 [redacted]

In addition to the analyses described in section 2.3 and 2.4, descriptive statistics for [redacted] will be presented.

The presentation of this analysis will be based on Version 1.0 of the Table, Figure and Listing Shells, Table 14.3.3.1.

2.8 References

Rom, D.M., *Strengthening Some Common Multiple Test Procedures for Discrete Data, Statistics in Medicine*, 11, 511-514 (1992)

2.9 Table Shells

Table 1 Analysis of 10-Metre Walk/Run Preferred Speed – Random Slopes and Intercepts Model
(Intent-To-Treat Analysis Set)

Parameter	Statistics	Tideglusib (N=XX)	Placebo (N=XX)
Mean at [redacted]	n	XX	XX
	Estimate	XX.X	XX.X
	SE	XX.XX	XX.XX
	95% CI	XX.X,XX.X	XX.X,XX.X
	Difference (Tideglusib - Placebo)	XX.X	
	SE	XX.XX	
	95% CI	XX.X,XX.X	
	p-value	X.XXXX	
Slope (change/day)	n	XX	XX
	Estimate	XX.X	XX.X
	SE	XX.XX	XX.XX
	95% CI	XX.X,XX.X	XX.X,XX.X
	Difference (Tideglusib - Placebo)	XX.X	
	SE	XX.XX	
	95% CI	XX.X,XX.X	
	p-value	X.XXXX	

N = the number of subjects in the analysis set. n = the number of subjects included in the analysis.
Analysis is a random slopes and intercepts model treatment by time interaction as fixed effects and intercept and time as random effects.
Source: Listing 16.2.6.7

PROGRAM NAME:T_XXX_01.SAS DATE OF RUN:XXXXXXXXX TIME OF RUN: XX:XX:XX DATE OF EXTRACTION:XXXXXXXXX

Programming note: Repeat for hand grip strength – dominant hand and creatine phosphokinase. Repeat these parameters for the per protocol set.

Table 2 Responder Analysis – Rom's Test
(Intent-To-Treat Analysis Set)

p-value	X.XXXX

[1] Rom's test is used to test the global null hypothesis of no treatment effect across multiple responder endpoints, versus the one-sided alternative of at least one effect in favour of tideglusib.

PROGRAM NAME:T_XXX_01.SAS DATE OF RUN:XXXXXXXXX TIME OF RUN: XX:XX:XX DATE OF EXTRACTION:XXXXXXXXX

Table 3 Summary of Correlations

(Intent-To-Treat Analysis Set)

Visit	Statistics	Tideglusib (N=XX)	Placebo (N=XX)	Overall (N=XX)
[redacted]	n	XX	XX	XX
	Pearson correlation coefficient	X.XX	X.XX	X.XX
	p-value	X.XXXX	X.XXXX	X.XXXX
	Spearman rank correlation coefficient	X.XX	X.XX	X.XX
	p-value	X.XXXX	X.XXXX	X.XXXX

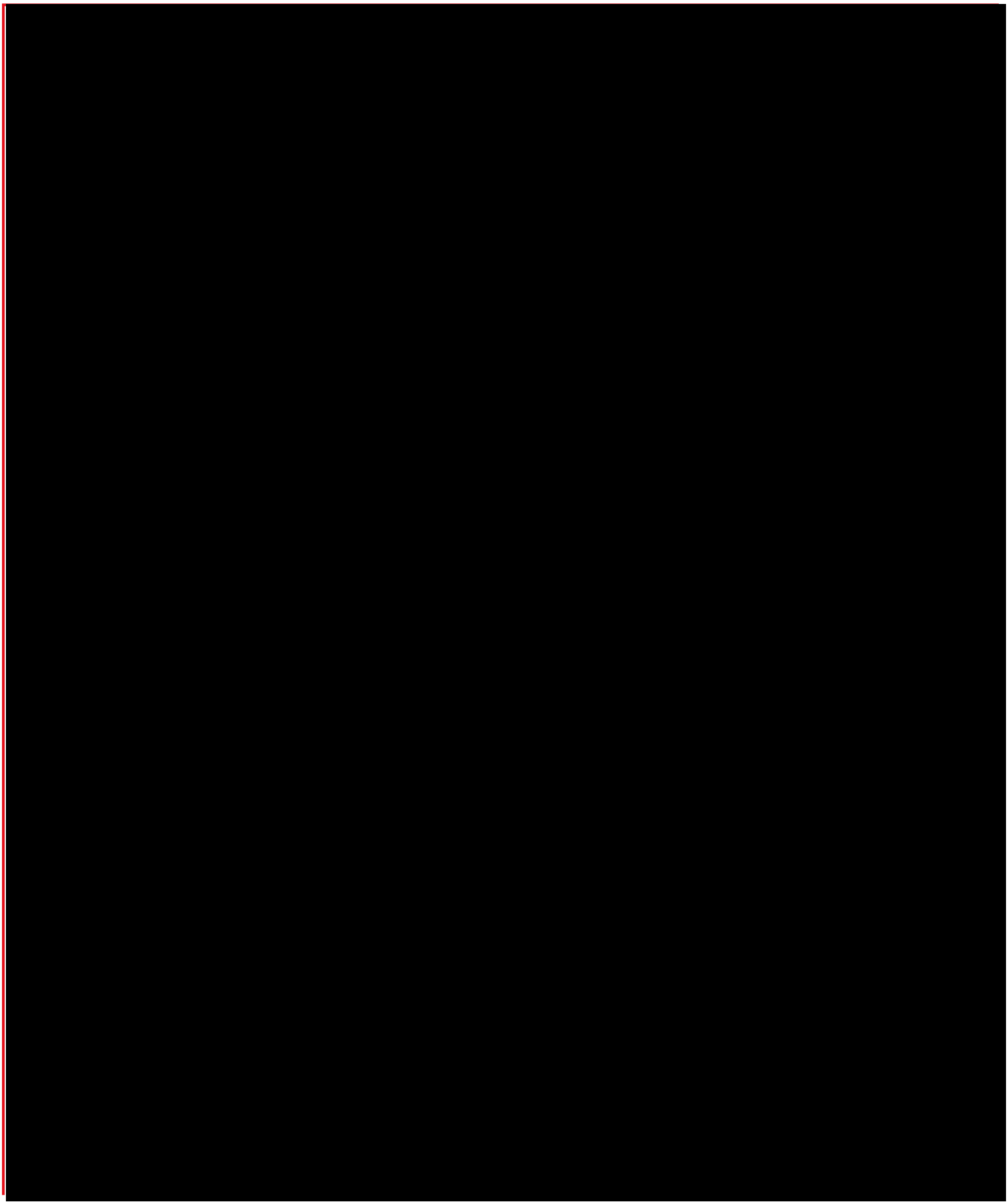
...

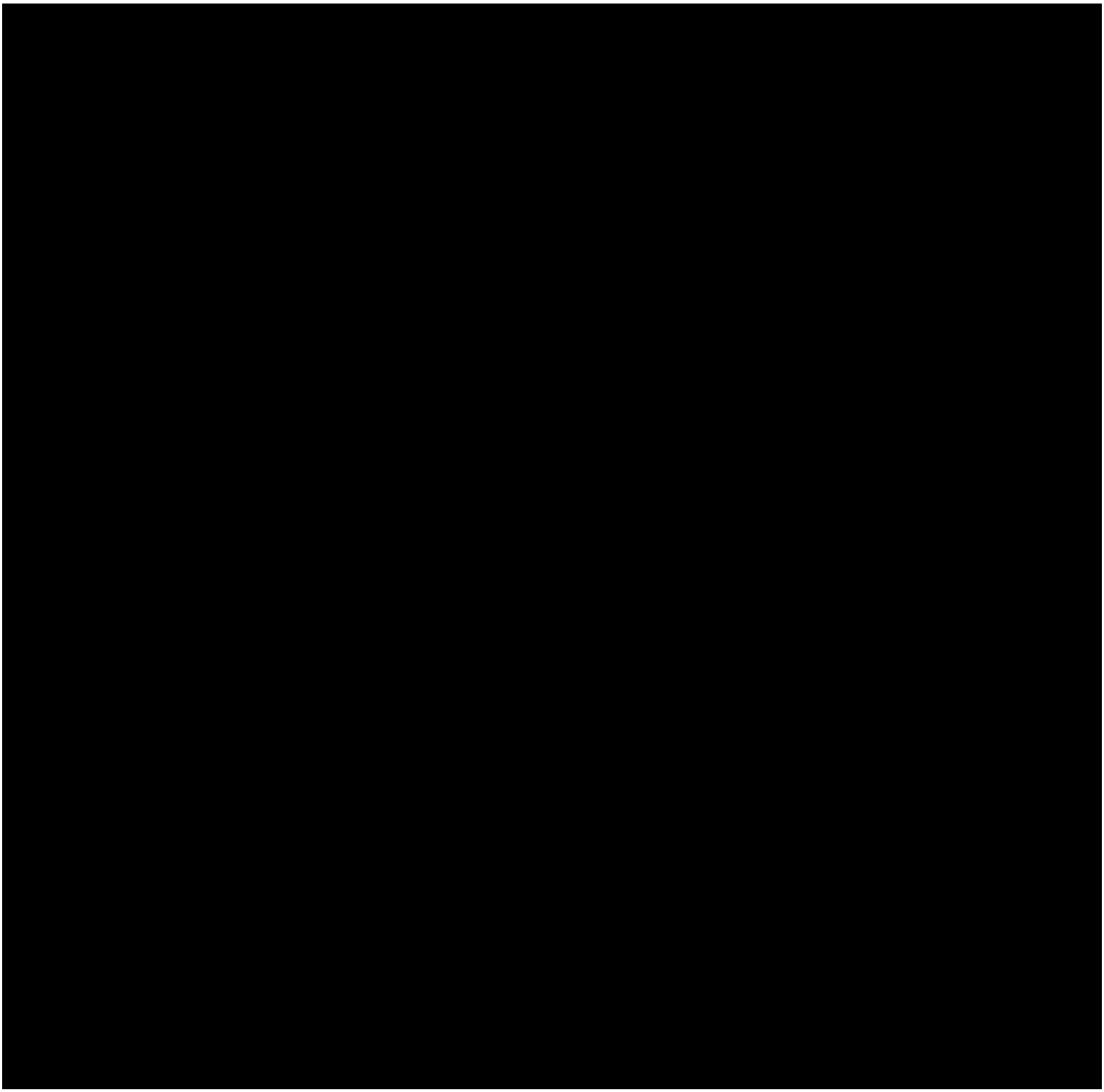
N = the number of subjects in the analysis set. n = the number of subjects with both assessments.

Source: Listing 16.2.6.1

PROGRAM NAME:T_XXX_01.SAS DATE OF RUN:XXXXXXXXX TIME OF RUN: XX:XX:XX DATE OF EXTRACTION:XXXXXXXXX

NOTES: ***Continue for all visits***





[redacted]

[redacted]

[redacted]