Appendix 1.9 Documentation of Statistical Methods

Statistical Analysis Plan dated 15 June 2021

Sponsor	GW Research Ltd
Protocol Title:	AN OPEN-LABEL EXTENSION TRIAL TO INVESTIGATE THE LONG-TERM SAFETY OF CANNABIDIOL ORAL SOLUTION (GWP42003-P, CBD-OS) IN PATIENTS WITH RETT SYNDROME
Protocol Number:	GWND19002
Premier Research PCN:	GWRE209830
EudraCT number:	2019-001605-24
ClinicalTrials.gov identifier	NCT04252586
Document Version:	FINAL
Document Date:	15-JUN-2021

Approvals

Role	Signatures	Date (dd-Mmm-yyyy)
Biostatistician	Print Name: PI PI Premier Research Sign Name: PI	15-Jun-2021 | 09:14:15 EDT
GW Research Ltd Representative	Print Name: PI Sign Name: PI	15-Jun-2021 | 07:38:28 PDT

Table of Contents

List of Tables

1. Overview

This statistical analysis plan (SAP) describes the planned analysis and reporting for GW protocol number GWND19002 (An open-label extension (OLE) trial to investigate the long-term safety of cannabidiol oral solution (GWP42003-P, CBD-OS) in patients with Rett Syndrome, dated 18 December 2019 version #2.0 and corresponding protocole annex #1, version 2 dated 11 November 2020. Reference materials for this statistical plan include the protocol and the accompanying sample data collection documents. GWND19002 is the extension trial of the GWND18064 trial which will be referred to here as lead-in trial. Operational aspects related to collection and timing of planned clinical assessments are not repeated in this SAP unless relevant to the planned analysis.

The study objectives listed below are according to latest Clinical Study Protocol and Annex. Due to potential impact of the COVID-19 pandemic to patients and site personnel the Sponsor made the decision to discontinue the randomized trial (GWND18064) as well as the extension trial. Specific inclusion of patients in this extension trial is covered by the Annex. As this extension trial is stopped prematurely some planned assessments as per protocol will not be performed, see Section 9.5 for complete description of the changes.

The structure and content of this SAP provides sufficient detail to meet the requirements identified by the Food and Drug Administration (FDA), European Medicines Agency (EMA), and International Conference on Harmonization (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use: Guidance on Statistical Principles in Clinical Trials[1].

The planned analyses identified in this SAP may be included in clinical study reports (CSRs), regulatory submissions, or future manuscripts. Also, post-hoc exploratory analyses not necessarily identified in this SAP may be performed to further examine study data. Any post-hoc or unplanned, exploratory analysis performed will be clearly identified as such in the final CSR.

The statistical plan described hereafter is an *a priori* plan. It will be finalized prior to any descriptive analysis of data pertaining to study GWND19002.

2. Study Objectives and Endpoints

2.1 Study Objectives

2.1.1 Primary Objective

The primary objective is to evaluate the long-term safety of GWP42003-P in patients with Rett syndrome (RTT).

2.1.2 Secondary Objectives

- To evaluate the effect of GWP42003-P in measures of disease severity:
  - Rett Syndrome Behaviour Questionnaire (RSBQ)
  - Clinical Global Impressions - Improvement (CGI-I)
  - Clinical Global Impressions - Severity (CGI-S)
  - 9-items Motor Behavioral Assessment (MBA-9)

- Children's Sleep Habits Questionnaire (CSHQ)

2.1.3 Exploratory Objectives

Exploratory Objectives:

- To evaluate the effect of GWP42003-P on caregiver and patient quality of life (QoL).
  - 36-item Short Form [SF-36] and Child Health Questionnaire Parent Form 50 [CHQ-PF50], respectively.
- To evaluate the effect of GWP42003-P on health utilization.
  - Hospital Services Use Questionnaire.
- Caregiver assessment of Rett symptoms (symptom diary).

2.2 Study Endpoints

2.2.1 Efficacy Endpoints

2.2.1.1 Primary Efficacy Endpoint

No primary efficacy endpoint in this open-label extension trial.

2.2.1.2 Secondary Efficacy Endpoints

The following will be assessed by evaluating changes relative to the prerandomization baseline of the GWND18064 lead-in trial:

- RSBQ
- CGI-I
- CGI-S.
- MBA-9.
- CSHQ.

Same analyses will be performed compared to OLE baseline.

2.2.1.3 Exploratory Efficacy Endpoint(s)

The following will be assessed by evaluating changes relative to the prerandomization baseline of the GWND18064 lead-in trial:

- Caregiver QoL questionnaire (SF-36).
- Patient QoL questionnaire (CHQ-PF50).
- Hospital Services Use Questionnaire.
- Caregiver assessment of Rett symptoms.

The same analyses will be performed compared to OLE baseline.

2.2.2 Safety Endpoints

The long-term safety profile of GWP42003-P will be assessed by evaluating changes in the following, relative to the prerandomization baseline of the GWND18064 lead-in trial:

- Adverse Events (AEs).

- Clinical laboratory parameters.
- Vital signs.
- Physical examination procedures.
- 12-lead electrocardiogram (ECG).
- Effects on menstruation cycles.
- Suicidality.
- Change in growth and development by measurement of height, weight, serum insulin-like growth factor-1 (IGF-1) levels and Tanner Staging (for patients aged ≥ 7 years, or earlier if clinically indicated by onset of menarche or other signs of precocious puberty).

The same analyses will be performed compared to OLE baseline.

3. Overall Study Design and Plan

3.1 Overall Design

This is a multicenter, OLE trial for patients with RTT who have completed the randomized, double blind, placebo-controlled trial (GWND18064).
Entry to this OLE trial is recommended to be on the same day as Visit 9 of the randomized controlled trial (RCT), GWND18064; however, patients may enter the OLE up to the RCT follow-up visit (Visit 11). For patients impacted by COVID-19, as per Annex, they are allowed to enroll into GWND19002 after the point of GWND18064 follow-up (Visit 11). For the ones who withdrew from GWND18064 due to COVID-19 pandemic containment measures or withdrew from GWND18064 due to sponsor administrative decision, they are allowed to enroll into GWND19002 at a later date, when appropriate.
All patients entering this OLE trial will begin dosing with 5 mg/kg/day GWP42003-P (2.5 mg/kg b.i.d.) on Day 1. Patients will be observed, and after 1 week, the dose may further be escalated, at the investigator's discretion, up to 15 mg/kg/day GWP42003-P (7.5 mg/kg b.i.d.) in weekly increments of 5 mg/kg/day (2.5 mg/kg b.i.d.). Patients should then remain on a stable dose of GWP42003-P for the duration of the maintenance period of the trial (up to 104 weeks), with the option for doses to be decreased or increased to a maximum dose of 20 mg/kg/day (10 mg/kg b.i.d.) based on clinical response and tolerability, as deemed necessary by the investigator, until the optimal dose is found. Patients whose dose has been decreased can have their dose increased again if tolerability improves. At the end of treatment (Visit 14 [Day 729]), the dose of GWP42003-P will be reduced over a 10-day taper period, and patients will enter the 4-week follow-up period.
If a patient permanently discontinues treatment at any point during the trial, GWP42003-P should be gradually reduced over 10 days (unless inadvisable due to an AE). The patient should attend a withdrawal visit (Visit 14) as soon as possible after the decision is made to permanently discontinue GWP42003-P. If applicable, the patient will taper GWP42003-P, attend the End of Taper visit (Visit 15), and then complete the 4-week follow-up period.

3.2 Sample Size and Power

There is no formal sample size calculation for this trial.
All patients with RTT who completed the randomized, double-blind, placebo-controlled trial (GWND18064) who wish to take GWP42003-P and who meet eligibility criteria or as per annex,

all patients who withdrew from GWND18064 due to COVID-19 pandemic containment measures or withdrew from GWND18064 due to sponsor administrative decision, are included in this trial.

3.3 Study Population

Same study population as the one defined at entry in lead-in trial:

- Female or male patients aged 2–18 years (inclusive), who weight at least 10 kg at entry in lead-in trial.
- At entry in lead-in trial, patients must have a clinical diagnosis of RTT (typical or atypical), defined according to RettSearch Consortium criteria. Patients must have a confirmed pathogenic genetic mutation of the *MECP2* gene.

Patients can be enrolled if they qualify for inclusion criteria and if they do not meet any exclusion criteria for GWND19002 protocol and annex. For complete list of inclusion/exclusion criteria, please consult clinical study protocol and annex.

3.4 Treatments Administered

The administered treatment is GWP42003-P oral solution. Mode of administration: to be taken orally twice daily (morning and evening) using the syringe(s) provided, preferentially with food i.e., within 30 minutes after the end of a meal and in line with the patients' normal feeding schedule and dietary habits. The time of IMP administration in relation to food should be kept consistent throughout the trial. Patients will take their first dose of IMP during GWND19002 trial at clinic visit Visit 1 and caregivers will be instructed how to measure and administer the IMP to the patient. All patients entering this OLE trial will begin dosing with 5 mg/kg/day GWP42003-P (2.5 mg/kg b.i.d.) on Day 1. Patients will be observed, and after 1 week the dose may escalate further, at the investigator's discretion, up to 15 mg/kg/day GWP42003-P (7.5 mg/kg b.i.d.) in weekly increments of 5 mg/kg/day (2.5 mg/kg b.i.d.).
Patients should then remain on a stable dose of GWP42003-P for the duration of the maintenance period of the trial (up to 104 weeks), with the option for doses to be decreased or increased if deemed necessary by the investigator, to a maximum dose of 20 mg/kg/day (10 mg/kg b.i.d.). Dose adjustments should be discussed with the GW medical monitor. The final decision regarding dose adjustments should be taken by the investigator.
At the end of treatment (Visit 14 [Day 729]), the dose of GWP42003-P will be reduced over a 10-day taper period, and patients will enter the 4-week follow-up period.
Patients discontinuing GWP42003-P treatment at any other time should undergo a 10-day taper period (unless continued dosing is inadvisable, e.g., due to an AE). The decision on whether to taper GWP42003-P or not will be left to the investigator's clinical judgment.

3.5 Method of Assigning Subjects to Treatment Groups

As this is an open-label study, with no assignment to treatment group, all patients will receive the GWP42003-P treatment.

3.6 Blinding and Unblinding

As this is an open-label study no blinding applicable.

3.7 Schedule of Assessments

A detailed schedule of events for the study is provided in Table 1 .

Table 1 Schedule of assessments

Visit Number	1	2[a]	3[a]	4	5	6	7	8	9	10	11	12	13	14b,c EoT/ Withdrawal	15 (end of taper)	16a (follow-up)
Day Number (Visit Window)	**1**	**8 (±3)**	**15 (± 3)**	**29 (± 7)**	**57 (± 7)**	**85 (± 7)**	**141 (± 7)**	**197 (± 7)**	**281 (± 7)**	**365 (± 7)**	**456 (± 7)**	**547 (± 7)**	**638 (± 7)**	**729 (± 7)**	**739 (+ 7)**	**767 (+ 7)**
Informed consent and assent[d]	X[e]															
Eligibility check	X[e]															
Concomitant medications		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
New medical history	X[e]															
Adverse events	X[e]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Menstruation question (where appropriate)								X		X				X		
Physical examination (including body weight)	X[f]			X		X		X	X	X		X		X	(X)	
Height								X		X		X		X		
ECG	X[f]			X		X		X		X		X		X	(X)	
Vital signs	X[g]			X	X	X	X	X	X	X		X		X	X	
Clinical laboratory blood sampling (hematology and biochemistry)[h]	X[f]			X	X	X		X	X	X		X		X	(X)	
Hepatic function monitoring panel[i]							X[i]				X[i]		X[i]			
Dipstick urinalysis (where possible)	X[f]							X		X		X		X	(X)	
Serum pregnancy test (if appropriate)								X		X		X		X		
Caregiver: RSBQ				X	X	X	X	X		X				X		
Caregiver: CSHQ						X		X		X				X		
Caregiver: Caregiver QoL questionnaire (SF-36)								X		X				X		

Visit Number		1	2[a]	3[a]	4	5	6	7	8	9	10	11	12	13	14b,c EoT/ Withdrawal	15 (end of taper)	16a (follow-up)
Day Number (Visit Window)		**1**	**8 (±3)**	**15 (± 3)**	**29 (± 7)**	**57 (± 7)**	**85 (± 7)**	**141 (± 7)**	**197 (± 7)**	**281 (± 7)**	**365 (± 7)**	**456 (± 7)**	**547 (± 7)**	**638 (± 7)**	**729 (± 7)**	**739 (+ 7)**	**767 (+ 7)**
	Patient QoL questionnaire (CHQ-PF50)								X		X				X		
	Hospital Services Use Questionnaire	X[g]			X	X	X	X	X		X				X		
	Tanner Staging (where appropriate)								X		X				X		
	Caregiver Assessment of Rett Symptoms				X	X	X	X	X		X				X		
MBA-9							X		X		X				X		
CGI-S					X	X	X	X	X		X				X		
CGI-I					X	X	X	X	X		X				X		
Suicidality assessment		X[g]			X	X	X	X	X	X	X		X		X	X	
GWP42003-P dispensing		X[e]			X	X	X	X	X	X	X	X	X	X	X		
GWP42003-P collection and compliance review					X	X	X	X	X	X	X	X	X	X	X	X	
Dosing schedule[j]		X	X	X	X	X	X	X	X	X	X	X	X	X	X		

(X) = if clinically indicated; EoT = end of treatment

a Visit to be conducted by telephone.

b To be performed to all patients completing or withdrawing from the trial. Patients who withdraw early should commence the 10-day GWP42003-P taper period, if possible.

c A safety follow-up Visit 4 weeks after last GWP42003-P dose is required for all patients who withdraw from the trial or complete the trial.

d Informed consent must be obtained prior to any trial-related procedures. In cases where the patient possesses adequate understanding, assent will be taken along with parent(s)/legal representative consent. The nominated caregiver will be asked to complete the quality of life questionnaires.

e Always required.

f Required only if OLE Visit 1 is > 28 days after RCT Visit 9 including enrollement after the point of RCT Visit 11 under annex.

g Required if OLE Visit 1 does not occur on the same day as RCT Visit 9.

h Determination of serum IGF-1 levels at Visits 1, 8, 10, and 14 only; IGF-1 laboratory results will remain blinded throughout the trial.

i Hepatic function monitoring is required following increases in GWP42003-P dose or introduction of medications that are known to impact liver function. If the concerned change does not occur within 1 month of a scheduled biochemistry assessment, the investigator should perform an additional hepatic

monitoring within 1 month of the change. Hepatic monitoring at Visit 7, 11 and 13 is required for patients taking concomitant valproic acid or whose dose exceeds 15 mg/kg/day.

j The dosing schedule is to be completed by the caregiver daily throughout the trial and reviewed at Visits 2 through 15.

4. Statistical Analysis and Reporting

Statistical analyses will be performed after the database is locked.

4.1 Introduction

Data processing, tabulation of descriptive statistics, and graphical representations will be performed primarily using SAS (release 9.4 or higher). If the use of other software is warranted, the final statistical methodology report will detail what software was used for what purposes.

Continuous (quantitative) variable summaries will include the number of patients (n) with non-missing values, the number of missing values, mean, standard deviation (SD), median, minimum, and maximum, unless otherwise specified.

Categorical (qualitative) variable summaries will include the frequency and percentage of patients who are in the particular category for each possible value. In general, the denominator for the percentage calculation will be based upon the total number of patients in the study population, unless otherwise specified.

The number of missing values will be calculated as difference between the total number of patients in the study population and the number of non-missing values.

The minimum and maximum will be reported with the same degree of precision (i.e. the same number of decimal places) as the observed data. Measures of location (mean and median) will be reported to 1 degree of precision more than the observed data and measures of spread (SD) will be reported to 2 degrees of precision more than the observed data.

Percentages will be presented to 1 decimal place, unless otherwise specified.

4.2 Interim Analysis, Data Review

4.2.1 Interim analysis

No formal interim analysis will be conducted.

4.2.2 Data review

No data review meeting will take place. After database lock, descriptive statistical outputs will be created.

5. Analysis Populations

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

Enrolled

- All patients from GWND18064 lead-in trial enrolled in GWND19002 trial.

Safety Analysis Set

- All patients who receive at least 1 dose of IMP in the GWND19002 trial will be included. Only patients for whom it has been confirmed that they did not take any IMP during the GWND19002 trial will be excluded from this safety analysis set. This analysis set will be used to report all statistical analysis.

Analysis sets will be identified during the conduct of statistical analysis after database lock.

6. General Issues for Statistical Analysis

All data collected in the study will be listed, ordered by site, patient number and, where applicable, chronological order of the assessment. Visit date is not needed to be included in the listings, day numbers will be included where appropriate.

Other derived variables (e.g. change from baseline values) that are calculated for analysis purposes or to aid interpretation of the data will be included in the data listings.

6.1 Statistical Definitions and Algorithms

6.1.1 Conventions for treatment/visit naming

In all tables, listings and figures, the treatment arms will be referred to and labelled as per Table 2.

Table 2 Study treatments

Endpoint	RCT Treatment	Actual Treatment	Protocol/Annex	Treatment Label (Table column/Figure legend)	Sorting order in listings, column order in tables
All	Double-Blind Treatment GWP42003-P different dosing	GWP42003-P, at different dosing	Enrolled under Protocol	Double Blind Treatment: GWP42003-P Enrolled Under Protocol	1
All	Double-Blind Treatment GWP42003-P different dosing	GWP42003-P, at different dosing	Enrolled under Annex	Double Blind Treatment: GWP42003-P Enrolled Under Annex	2
All	Double-Blind Treatment Placebo different dosing	GWP42003-P, at different dosing	Enrolled under Protocol	Double Blind Treatment: Placebo Enrolled Under Protocol	3
All	Double-Blind Treatment Placebo different dosing	GWP42003-P, at different dosing	Enrolled under Annex	Double Blind Treatment: Placebo Enrolled Under Annex	4

In all tables, listings and figures, the study visits will be referred to and labelled as per Table 3.

Table 3 Study Visits

Actual Visit	Visit Label
Baseline (last record before 1st IMP administration during lead-in trial)	RCT Baseline
Baseline (Assessment at extension Visit 1)	OLE Baseline
Visit 1: Day 1	Day 1
Visit 2: Day 8 Telephone Call	Day 8 TC
Visit 3: Day 15 Telephone Call	Day 15 TC
Visit 4: Day 29	Day 29
Visit 5: Day 57	Day 57
Visit 6: Day 85	Day 85
Visit 7: Day 141	Day 141
Visit 8: Day 197	Day 197
Visit 9: Day 281	Day 281
Visit 10: Day 365	Day 365
Visit 12: Day 547	Day 547
Visit 14: Day 729 or withdrawal	End of Treatment
Visit 15: Day 739 (end of taper)	End of Taper
Visit 16: Day 767	Follow-up

No derivation will be used for visits; data will be tabulated and listed according to the visit it was recorded under in eCRF.

6.1.2 Baseline

For all endpoints, two baselines will be defined:

- the last non-missing observation recorded prior to the first drug administration during GWND18064 lead-in trial will be used as the RCT baseline
- The value reported at Visit 1 during OLE study will be used as the OLE baseline

6.1.3 Last Visit

The last visit for endpoints assessed at clinic visits is defined as the last scheduled visit at which a patient's last evaluation (up to Visit 15/end of treatment visit/any other last visit) is performed.

6.1.4 Day Numbering

The first day of treatment (Day 1) will be the date of Visit 1.

Any days prior to Day 1 will be numbered relative to this day and calculated as:

Date − (Date of Day 1)

to give Day −1, −2, −3 etc.

Any days post Day 1 will be calculated as:

1 + Date − (Date of Day 1)

6.1.5 Multiple Comparisons

No multiple comparisons are planned as no inferential analysis are planned.

6.1.6 Handling of Dropouts or Missing Data

Every effort has been made to minimize missing data for this trial. The patient reported outcomes will be completed by the patient's primary caregiver on electronic devices. These assessments have been set up in such a way that the caregiver cannot proceed to the next question if they have not entered a response for the current question. In addition, on the off chance that the device does not function correctly, paper alternatives of the assessments are available.

6.1.6.1 Handling of Missing Data for Efficacy Endpoint (Primary, Secondary and Exploratory)

If a questionnaire has subscales then the missing items in a subscale can be imputed as the mean of the non-missing items in that subscale, provided that at least half of the items in the subscale were answered. If more than half the items in the subscale are missing then the missing items will be imputed as LOCF.

If a questionnaire does not have subscales (eg MBA-9, CSHQ single item scale) the missing items will be imputed using LOCF.

6.1.6.1.1 SF-36

The SF-36 assessment will be scored by the QualityMetric Health Outcomes Scoring Software 5.0 (Maruish, 2011) which assumes that the response to a missing item in a particular scale is the same as the mean of the responses to the scale's answered items. This approach cannot be used to estimate item responses on the physical functioning scale due to the hierarchical nature of the items included in this scale. If items on the physical functioning scale are missing, the QualityMetric Health Outcomes Scoring Software v5.0 estimates the PF score using item response theory. If any SF-36 is logged as completed by an alternative caregiver, it will be counted as missing.

6.1.6.1.2 CHQ-PF50

Appendix 5 describes the methods of handling missing data for the CHQ-PF50 health scales.

6.1.6.1.3 Adverse Events

Missing and/or incomplete dates for AEs will be imputed in a manner resulting in the earliest onset or the longest duration during the treatment period, while ensuring that the start date is not after the stop date, i.e. missing start dates will be imputed as the first day of the month/first month of the year and missing stop dates will be imputed as the last day of the month/last day of the year, while ensuring that the start date is not after the stop date. Upper and lower limits of imputation will be date of Visit 1 and last visit date as defined in section 6.1.3, and events occurring on date of Visit 1 will be assumed post-IMP. Stop dates will not be imputed if the AE is ongoing.

Imputed dates will be used for calculation of time to first onset and time to AE resolution, they will not be presented in AE listings.

A worst-case approach will be followed in the event of missing severity or causality data. If the

severity is missing, 'Severe' will be imputed. If causality data is missing for a treatment emergent AE (TEAE), 'Yes' will be imputed for the question 'Plausible relationship to study medication'.

6.1.6.1.4 Concomitant Medication

Missing concomitant medication dates will be handled in a similar fashion as described for AEs in Section 6.1.6.1.3.

6.1.7 Pooling of Sites

As there are expected to be relatively few patients per site, the site will not be taken into account in the analyses. Therefore the question of pooling of sites does not arise.

6.1.8 Derived Variables

For derivations of the efficacy variables, see Section 8. The following additional derived and computed variables have been initially identified as important for the analyses to be performed for this trial.

In case additional derived variables may be required: the SAP will not be amended for additional variables that are not related to the primary or secondary endpoints. Any additional derived or computed variables will be identified and documented in the SAS programs that create the analysis files. If the SAP is not amended, further derivations related to the primary and secondary endpoints will be described in the CSR.

6.1.8.1 Adverse events

Any adverse events occuring between last dose of RCT and OLE visit 1 will not be considered Treatment Emergent Adverse events (Aes occuring during study break). All other adverse events will be considered as treatment-emergent adverse events.

A TEAE will be considered treatment-related if the plausibility relationship to IMP is recorded on the eCRF as 'yes'. If the data on plausibility relationship to IMP is missing then the TEAE will be considered treatment-related.

A TEAE will be considered leading to permanent discontinuation of IMP (leading to withdrawal) if the action taken with IMP is recorded on the eCRF as 'trial medication stopped' or the outcome is recorded on the eCRF as 'patient died'.

A treatment-related TEAEs leading to withdrawal is a TEAE leading to permanent IMP discontinuation and considered treatment-related as per above definition.

A TEAE will be considered leading to IMP temporary discontinuation if the action taken with IMP is recorded on the eCRF as 'trial medication interrupted'.

A TEAE will be considered leading to permanent IMP dose reduction excluding permanent discontinuation if the action taken with IMP is recorded on the eCRF as 'dose reduced'.

A TEAE will be considered fatal if the outcome is recorded on the eCRF as 'patient died'.

The time to first onset of AE will be calculated for TEAEs as:

Start date of AE – Day 1 + 1

The time to AE resolution will be calculated for TEAEs as:

Stop date of AE – Start date of AE + 1

6.1.8.2 Concomitant medications

A concomitant medication is defined as any medication that was administered during the treatment period (from Visit 1 of OLE trial). This includes medications that started before the treatment period and continued while on treatment and medications that started during the treatment period.

6.1.8.3 Age

Age will be calculated as:

(Date of Visit 1 – date of birth) ÷ 365.25.

6.1.8.4 Exposure

The total number of dosing days in the treatment period will be calculated as:

(Date of last IMP dose from Study Outcome eCRF form – Date of Day 1) + 1

For Day 1 definition see section 6.1.4

6.1.9 Data Adjustments/Handling/Conventions

All collected data will be presented in listings. Data not subject to analysis according to this plan will not appear in any tables or graphs but will be included only in the data listings.

7. Study Patients and Demographics

7.1 Disposition of Patients and Withdrawals

Patient disposition, by site, by country and overall, will be summarized overall using standard summary statistics. Patients enrolled under Protocol / Annex will also be described. The number of patients screened, number of screen failures and number enrolled will be included.

A screen failure disposition table will be presented, including number of patients screened, number failing screening, number enrolled and the reasons for failing screening.

Patient disposition, including patients treated, patients who completed the treatment period and the taper phase, patients discontinued (including reason for discontinuation) from the treatment, and taper phases will be summarized by absolute counts (n) and percentages (%). Number and percentages of patients completing each study visit will be described. A further table split by site, and by country will be produced, showing number of patients enrolled, withdrawn at each site or in each country.

7.2 Protocol Deviations

Protocol deviations will be identified and classified.

Major protocol deviations will be summarized by type of violations for the enrolled population.

Protocol deviations will be listed. Protocol deviations due to COVID-19 will be listed separately.

7.3 Demographics and Other Baseline Characteristics

The following demographic data will be summarized overall for the safety analysis sets:

- Age (years);
- Age group (2-5 years, 6-12 years and 13-19 years);
- Sex;
- Race;
- Ethnicity
- Country;
- Region (United States, Rest of the World);

7.3.1.1 History of RTT

The following history of RTT data will be summarized by treatment group and overall for the safety analysis set:

- Type of diagnosis (Typical, Atypical).
- Main and supportive criteria for RTT diagnosis.
- Age at diagnosis (months)
- Time since last known loss of hand use or verbal language or gross motor regression (months).
- Information about the *MECP2* mutation (whether a known mutation is documented and if available, the type of mutation).
- Communication methods used by the child.
- Types of seizures that are ongoing.
- Status epilepticus.
- Whether the child lives at home or at a care home/institution.
- Most common feeding method.
- Planned method for IMP administration.

7.3.2 Medical and Surgical History and Current Medical Conditions

All conditions and diagnoses on the 'medical and surgical history' eCRF page will be coded using Version 21.1 of the Medical Dictionary for Regulatory Activities (MedDRA v21.1) or later version.

Only new medical history not already reported in the lead-in study will be described there.

The number of patients with relevant or significant new medical or surgical history by system organ class, and preferred term, will be summarized by absolute counts (n) and percentages (%). Percentages will be calculated based on the number of patients in the analysis population.

7.4 Exposure and Compliance

IMP is to be administered twice daily (morning and evening). The first dose will be taken on Day 1 (Visit 1). The date of final dose in the treatment period will be recorded on the eCRF under DS

form. The date of final dose, for patients who enter the taper period, will be recorded on the eCRF at the end of taper visit under DS2 form. Total number of dosing days will be reported.

8. Efficacy Analysis

The efficacy analyses will use the safety analysis set. No formal hypothesis testing will be carried out for this study. Statistical analysis will be descriptive.

8.1 Rett Syndrome Behaviour Questionnaire (RSBQ)

The RSBQ is a caregiver-completed questionnaire that measures the frequency of current disease characteristics (45 items) that may or may not apply to the patient. Each item is rated on a 3-point numerical scale, where all items except for item 31 ("Uses eye gaze to convey feelings, needs and wishes") is scored as follows: 0 indicating an item that is 'not true as far as you know', 1 indicating an item is 'somewhat or sometimes true', and 2 indicating an item that is 'very true or often true'. Item 31 is to be reverse scored, where 0 indicates that this item is 'very true or often true', 1 indicating that this item is 'somewhat or sometimes true', and 2 indicating that this item is 'not true as far as you know'. The total maximum score is 90 and higher total scores represent greater severity.

The RSBQ score and its change from baseline (RCT Baseline and OLE Baseline) will be summarized on a numerical scale by visit.

8.1.1 RSBQ Subscales

The RSBQ includes 8 subscales: general mood, breathing problems, hand behaviors, face movements, body rocking/expressionless face, night-time behaviors, anxiety/fear, and walking/standing. The items included in each subscale are presented in Table 4.

Table 4 RSBQ Subscales

Subscale	Items Included
General Mood	2, 14, 15, 16, 22, 29, 30, 36
Breathing Problems	1, 5, 6, 19, 25
Hand Behaviors	18, 20, 21, 24, 35, 43
Face Movements	4, 28, 32, 34
Body Rocking/Expressionless Face	12, 17, 31*, 33, 40, 41
Night-time Behaviors	13, 37, 42
Anxiety/Fear	7, 9, 10, 38
Walking/Standing	23, 39

* Item 31 is to be reverse scored as described in section 5.5.2.

The subscale scores are calculated by summing the scores of the items in each subscale.

Subscale scores and their change from baseline will be summarized by visit for each subscale.

8.2 Clinical Global Impressions - Improvement (CGI-I)

The CGI-I is a 7-point scale that requires the clinician to assess how much the patient's illness has improved or worsened relative to a baseline state at the beginning of the intervention. This is rated as: 1 = very much improved; 2 = much improved; 3 = minimally improved; 4 = no change; 5 =

minimally worse; 6 = much worse; or 7 = very much worse.

The CGI-I score will be summarized on both a categorical scale and a numerical scale by visit.

8.3 Caregiver Global Impression of Severity (CGI-S)

The CGI-S is a 7-point scale that requires the clinician to rate the severity of the patient's illness at the time of assessment, relative to the clinician's experience with patients who have the same diagnosis. Considering total clinical experience, a patient will be assessed on severity of illness at the time of rating. This is rated as: 1 = normal, not at all ill; 2 = borderline ill; 3 = mildly ill; 4 = moderately ill; 5 = markedly ill; 6 = severely ill; or 7 = extremely ill.

CGI-S scores and their change from baseline (RCT baseline and OLE baseline) will be summarized by treatment and visit. Frequency table will be created for CGI-S categories by treatment and visit.

8.4 9-items Motor Behavioral Assessment (MBA-9)

The MBA-9 scale is completed by the investigator and addresses core symptoms of RTT. MBA-9 was derived from the full MBA scale (37 RTT symptoms) by selecting the items that are deemed to be amenable to change and that reflect areas of meaningful clinical change. The severity of current symptoms is rated on a 5-point numerical scale; 0 = normal or never; 1 = mild or rare; 2 = moderate or occasional; 3 = marked or frequent; 4 = very severe or constant. The MBA-9 score is calculated by summing the scores of the individual items. The maximum score is 36 and higher scores represent greater severity.

MBA-9 scores and their change from baseline (RCT baseline and OLE baseline) will be summarized by visit.

8.4.1 Children's Sleep Habit Questionnaire (CSHQ)

The CSHQ includes a total sleep disturbance score and 8 sleep subscales: bedtime resistance, sleep onset delay, sleep duration, sleep anxiety, night wakings, parasomnias, sleep disordered breathing and daytime sleepiness. The items included in each subscale are presented in Table 5.

Table 4 CSHQ Subscales

Subscale	Items Included
Bedtime resistance	1*, 3*, 4, 5, 6, 8
Sleep onset delay	2*
Sleep duration	9, 10*, 11*
Sleep anxiety	5, 7, 8, 21
Night wakings	16, 24, 25
Parasomnias	12, 13, 14, 15, 17, 22, 23
Sleep disordered breathing	18, 19, 20
Daytime sleepiness	26*, 27, 28, 29, 30, 31, 32, 33

* These items are to be reverse scored.

The responses to each item of the CSHQ are to be scored as Usually=3, Sometimes=2, Rarely=1, except for the items marked with * in Table 5, which are considered to be "desirable" sleep behaviors and are therefore reverse scored. Items 32 and 33 are scored 1 for "not sleepy", 2 for

"very sleepy", 3 for "falls asleep".

CSHQ scores and CSHQ subscale scores along with their change from baseline (RCT baseline and OLE baseline) will be summarized by visit.

8.5 Exploratory Efficacy Endpoints

8.5.1 Caregiver Quality of Life Questionnaire

The caregiver's health-related quality of life will be assessed using the 36-item short form (SF-36). The SF-36 measures 8 domains of health-related quality of life: 1) limitations in physical activities because of health problems; 2) limitations in social activities because of physical or emotional problems; 3) limitations in usual role activities because of physical health problems; 4) bodily pain; 5) general mental health (psychosocial distress and well-being); 6) limitations in usual role activities because of emotional problems; 7) vitality (energy and fatigue); and 8) general health perceptions. These domains are used to calculate 2 composite scores, the physical health composite score (PCS) and the mental health composite score (MCS).

The items included in each subscale are presented in Table 6.

Table 5 SF-36 Subscales

Subscale	Items Included
Physical functioning	3 (a to j)
Role-physical	4 (a to d)
Bodily pain	7 and 8
General health	1 and 11 (a to d)
Vitality	9 (a, e, g and i)
Social functioning	6 and 10
Role-emotional	5 (a to c)
Mental health	9 (b, c, d, f and h)

The raw SF-36 data will be sent to Optum via Rave Web Services so that the subscales and composite scales can be scored by the QualityMetric Health Outcomes™ Scoring Software 5.0. The scores will then be uploaded into the electronic database.

The scoring of the SF-36 questionnaire via the QualityMetric Health Outcomes™ Scoring Software 5.0 is based on norm-based scoring which ensures that the scores for the subscales have a mean of 50 and a standard deviation of 10, based on the 2009 US general population (Maruish, 2011). A higher score is indicative of a better health state.

The steps that the software follows to score the scales and composite scales are summarized below:

1. Some of the items are recoded to ensure that a higher score corresponds to a better health state for all items in the SF-36.
2. The raw scores for the health domain scales are calculated as the sum of the final response values for the items included in each scale.
3. Transform the score to a 0-100 scale score:

   Transformed scale score = (Actual raw score - Lowest possible raw score)/Possible raw score range x 100

4. Transform to z-scores ~ Normal(0,1) according to the 2009 US general population norms.
5. Calculate T scores for each domain = *T* score = 50 + (domain *z*-score x 10). T scores calculated to aid interpretation.
6. Finally calculate MCS and PCS using a 3 step procedure:

   First, the eight health domain scales are standardized using means and standard deviations from the 2009 U.S. general population. Second, these standardized scores are aggregated using weights (factor score coefficients) from the 1990 U.S. general population. Third, aggregate PCS and MCS scores are standardized using a linear *T*score transformation with a mean of 50 and a standard deviation of 10.

SF-36 composite scores and SF-36 subscale scores along with their change from baseline (RCT baseline and OLE Baseline) will be summarized by visit.

8.5.2 Patient Quality of Life Questionnaire

This instrument is a well-validated general QoL measure in pediatric populations with chronic illness. It measures QoL of the child and the family by parent or child report. The caregiver will be asked to complete the questionnaire on behalf of the patient. The CHQ-PF50 covers multidimensional health concepts including Physical Functioning, Role/Social Limitations–Emotional/Behavioral, Role/Social Limitations– Physical, Behavior, Mental Health, Self-Esteem, General Health , Bodily Pain, Family Activities, Parent Impact–Time, Parent Impact–Emotional, and Family Cohesion. Scores for specific items and subscales, as well as a standardized physical summary (PhS) score and a standardized psychosocial summary (PsS) score, can be calculated from the CHQ-PF50 questionnaire. Scores are based on a 0 to 100 scale and higher scores indicate better quality of life.

Table 7 presents the items included in each health concept.

Table 6 CHQ-PF50 Health Concepts

Health Concept	Items Included
Global health item	1.1
Physical functioning	2.1 (a to f)
Role/social limitations due to emotional/behavioral difficulties	3.1 (a to c)
Role/social limitations due to physical health	3.2 (a to b)
Bodily pain and discomfort	4.1 and 4.2
Behavior	5.1 (a to e) and 5.2
Global behavior item	5.2
Mental health	6.1 (a to e)
Self esteem	7.1 (a to f)
General health perceptions	8.1 (a to e) and 1.1
Change in health item	8.2
Emotional impact on parent	9.1 (a to c)
Time impact on parent	9.2 (a to c)
Family activities	9.3 (a to f)
Family cohesion item	9.4

Details regarding the scoring of the health concepts of CHQ-PF50 and of the PhS and PsS can be found in Appendix 5. All of the health concepts and scales included in CHQ-PF50, except for the

change in health item are continuous.

CHQ-PF50 standardized scores (as per Appendix 5) CHQ-PF50 subscale scores, PhS and PsS scores along with their change from baseline (RCT Baseline and OLE Baseline) will be summarized by visit.

8.5.3 Hospital Services Use Questionnaire

The Hospital Services Use Questionnaire captures the frequency of hospital visits and patient hospitalizations. The responses to this questionnaire will be listed only.

8.5.4 Caregiver Assessment of Rett Symptoms

The Caregiver Assessment of Rett Symptoms is a symptom diary, including 12 items, each scored between 0 and 10, where a higher score for all items except items 1 and 9 represents a greater severity. Items 1 and 9 will be reverse-scored and a total score will be calculated as the sum of items 1 to 10, 11a and 12a. For items 11 and 12, if 11 is 'No' then 11a should be scored 10 and if 12 is 'No' then 12a should be scored 0. A visit score will be derived for each item and for the total score as the mean of the scores since the previous visit.

The total scores and the scores for each item of the Caregiver Assessment of Rett Symptoms along with change from baseline (RCT baseline and OLE baseline) will be summarized by visit.

9. Safety and Tolerability Analysis

9.1 Adverse Events

All reported AEs will be classified by system organ class (SOC), and preferred term (PT) and lower level term using Version 21.1 of MedDRA or higher.

The following summaries will be generated (counts are by patient unless specified otherwise):

Overall summary of TEAEs, including following description:

- Summary of TEAEs.
- Summary of treatment-related TEAEs.
- Summary of TEAEs leading to permanent discontinuation of IMP.
- Summary of serious TEAEs.
- Summary of treatment-related serious TEAEs.

Following summaries will be presented by SOC and PT:

- Summary of TEAEs
- Summary of treatment-related TEAEs
- Summary of TEAEs by maximal severity.
- Summary of serious TEAEs
- Summary of non-serious TEAEs.
- Summary of treatment-related serious TEAEs.
- Summary of TEAEs leading to permanent discontinuation of IMP.
- Summary of treatment-related TEAEs leading to permanent discontinuation of IMP.

- Summary of TEAEs leading to IMP dose reduction excluding permanent discontinuation (by resolution and overall).
- Summary of treatment-related TEAEs leading to IMP dose reduction excluding permanent discontinuation (by resolution and overall).
- Summary of TEAEs leading to IMP interrupted (by resolution and overall).
- Summary of treatment-related TEAEs leading to IMP interrupted (by resolution and overall).
- Summary of fatal TEAEs.
- Summary of TEAEs by time of first onset of AE.
- Summary of TEAEs by time to AE resolution.
- Summary of TEAEs reported in ≥ 2% of patients (after rounding)
- List of patients experiencing TEAEs by SOC and preferred term.

For the summary of TEAEs by maximal severity, for each patient, the worst severity recorded by PT, SOC and overall will be used for summary purposes. If severity is missing, the worst case (severe) will be assumed.

For summaries by resolution, AEs with an outcome of 'recovered' or 'recovered with sequelae' will be summarized as 'Resolved' and AEs with an outcome of 'continuing', 'patient died' or those with a missing outcome will be summarized as 'Not resolved'.

For the summary of TEAEs by time of first onset of AE, data will be summarized under the following categories:

- Month 1 (Day 1–29).
- Month 2 (Day 30–59).
- Month 3 (Day 60–89).
- Month 4-6 (Day 90–179).
- Month 7-9 (Day 180–269).
- Month 10-12 (Day 270–359).
- >1 year (> Day 359).

If patients have multiple occurrences of an AE then the AE will be counted once for the first occurrence only. Percentages will be based on the number of patients in the safety analysis set who have a visit or follow-up call within each time period above. In case there is no planned visits in the interval, percentages will be based on the number of patients in the safety analysis set.

For the summary of TEAEs by time to AE resolution, data will be summarized under the following categories:

- 1 week (≤7 days).
- 2 weeks (8–14 days).
- 3 weeks (15–21 days).
- 4 weeks (22–28 days).
- >4 weeks (>28 days).
- Ongoing (for AEs not resolved).

If patients have multiple occurrences of an AE then the AE will be counted once for the occurrence with the longest time to AE resolution. However, if any of the AEs are not resolved then the AE will be counted once within the 'Ongoing' category.

The start and stop day of the AE relative to Day 1 will be calculated as per Section 6.1.4.

All TEAEs will be listed. Listings will include the start and stop day of the AE, and limited demographic information about the patient (age, sex, race and weight at screening). A separate listing will be provided for serious AEs and events of special interest (see APPENDIX).

9.2 Clinical Laboratory Evaluation

9.2.1 Hematology and Biochemistry

Hematology and biochemistry safety parameters are measured at Visit 1, Visit 4, Visit 5, Visit 6, Visit 8, Visit 9, Visit 10, Visit 12 and Visit 14 (end of treatment).

Summaries will be presented for each laboratory parameter at each visit. Change from baseline (RCT and OLE baseline) and percent change from baseline to each post-baseline visit will also be presented.

If values for any of the parameters are below or above the limit of quantification of the assay (BLQ or ALQ), then they will be included in the summary tables at the BLQ or ALQ thresholds.

Where laboratory samples are repeated, the baseline value is defined as the final recorded value prior to the first dose of IMP during the lead-in trial or prior to Visit 1 in OLE trial.

Shift tables for hematology and biochemistry parameters will be produced, based upon normal ranges and GW toxicity limits (See Appendix 3), to determine the categorical shifts from each baseline (RCT and OLE) to each post-baseline visit. Values will be categorized as 'Normal', 'Low' or 'High' based on normal ranges and 'Toxically Low', 'Toxically Normal' or 'Toxically High' based on GW toxicity limits.

For eGFR, results will be assigned to the following grades:

- Normal: >60 ml/min/1.73 m^2
- Grade 1: 60 ml/min/1.73 m^2
- Grade 2: ≥30 and <60 ml/min/1.73 m^2
- Grade 3: ≥15 and <30 ml/min/1.73 m^2
- Grade 4: <15 ml/min/1.73 m^2

A separate shift table will be produced for eGFR based upon the above grades to determine the categorical shifts from baseline to each post-baseline visit.

An additional table will be produced, summarizing the number of patients meeting the following criteria:

- Alanine aminotransferase (ALT) > 1×ULN at baseline
- Aspartate aminotransferase (AST) > 1×ULN at baseline
- AT > 1×ULN at baseline
- Treatment-emergent ALT > 3×ULN, > 5×ULN and > 8×ULN
- Treatment-emergent AST > 3×ULN, > 5×ULN and > 8×ULN
- Treatment-emergent AT > 3×ULN, > 5×ULN and > 8×ULN
- Treatment-emergent AT > 3×ULN and either bilirubin > 2×ULN or INR > 1.5

where AT is AST or ALT, and treatment emergent is defined as criteria not met at baseline, but

met at any time post-baseline. The above will be summarized overall and for the following subgroups:

- Valproic acid use (as a concomitant medication) (Yes, No).
- Clobazam use (as a concomitant medication) (Yes, No).
- Valproic acid use and Clobazam use (as concomitant medications) (Yes/Yes, Yes/No, No/Yes, No/No).
  (A medication will be considered concomitant if it has a start date on or after the first dose of IMP or if it was started prior to the first dose of IMP and was ongoing.)
- Patients taking 3 or more current AEDs.
- Patients taking 4 or more current AEDs.

All laboratory data will be listed; listings will include limited demographic information about the patient (age, sex, race and weight at baseline). Abnormal laboratory values will be listed separately. A further listing will be created for the laboratory reference ranges and toxicity limits. A listing of liver parameters (ALT, AST, bilirubin and INR) will be produced that includes the baseline result, result at the particular visit, change from baseline, upper limit of normal (ULN) value, ratio of result to baseline and ratio of result to ULN.

9.2.2 Urinalysis

Urinalysis is assessed using dipsticks at Visit 8, Visit 10, Visit 12, and Visit 14 (end of treatment).

Urinalysis results will be listed only.

9.2.3 Pregnancy Test

Serum pregnancy test results will be summarized by visit.

9.3 Vital Signs, Other Physical Findings and Other Safety Data

9.3.1 Vital Signs

Vital signs (systolic blood pressure, diastolic blood pressure and pulse rate) are measured at Visit 1 , Visit 4, Visit 5, Visit 6, Visit 7, Visit 8, Visit 9 , Visit 10, Visit 12, Visit 14 (end of treatment) and Visit 15 (end of taper).

Summaries will be presented for each vital sign parameter at each visit. Change from baseline RCT and OLE baseline) to each post-baseline visit will also be presented.

A separate table will be produced, by visit, presenting the incidence of patients with vital signs indicative of a medical condition at Visit 1 and indicative of an AE after Visit 1.

Based on the criteria presented in Appendix 3, potentially clinically significant changes from baseline (RCT and OLE baseline) in vital signs measurements and other defined flagged values will be identified at each visit. The number of patients with a potentially clinically significant change from baseline will be summarized by parameter, and visit. The number of patients with at least one post-baseline flagged vital sign parameter value will be summarized by parameter, and flagged criteria.

9.3.2 Electrocardiogram

An ECG will be performed at Visit 1, Visit 4, Visit 6, Visit 8, Visit 10, Visit 12 and Visit 14 (end of treatment).

Summaries will be presented for mean heart rate, RR interval, PR interval, QRS duration, QT interval, QTcB and QTcF, at each visit. Change from baseline (RCT and OLE baseline) to each post-baseline visit will also be presented.

A separate table will be produced, by visit, presenting the incidence of patients

Based on the criteria presented in Appendix 3, defined flagged values will be identified at each visit. The number of patients with at least one post-baseline flagged ECG parameter value will be summarized by parameter, and flagged criteria.

9.3.3 Physical Examination

A physical examination which includes body weight measurements will be performed at Visit 1, Visit 4, Visit 6, Visit 8, Visit 9, Visit 10, Visit 12, and Visit 14 (end of treatment). Height will be collected as part of the physical examination at Visit 8, Visit 10, Visit 12, Visit 14 (end of treatment).

Any relevant findings at screening are included as part of the patient's medical history. Any changes seen after screening that are indicative of an AE are to be recorded as such on the AE form and included as part of the AE summaries.

Weight along with weight change and height data will be summarized by visit where applicable.

9.3.4 Suicidality Assessment

The profound cognitive impairment of RTT patients is such that the Children's Columbia-Suicide Severity Rating Scale is not considered appropriate in this trial. Instead, suicidality will be assessed by the investigator via a clinical interview with the caregiver.

Responses to the suicidality assessment will be summarized by visit.

9.3.5 Growth and Development

IGF-1 levels will be analyzed as part of the clinical laboratory testing. IGF-1 levels will be summarized on a continuous scale, including change from baseline.

Change from baseline (RCT and OLE baseline) to the end of treatment visit for IGF-1 levels will also be plotted against the Tanner Stages, weight, and height recorded at baseline.

The pubic hair growth (both sexes), genital (males only) and breast (females only) development of all adolescent patients (i.e., $\geq$ 7 years of age at the time of signing the informed consent form, or earlier if clinically indicated by onset of menarche or other signs of precocious puberty) will be assessed using Tanner Staging. The assessment can either be performed by examination during the study visit or the appropriate Tanner Stage can be indicated by the caregiver, with reference to the chart provided.

Patients will be examined at Visit 8, Visit 10 and Visit 14 (end of treatment). Once a patient

reaches a score of V (i.e., 5) the examination need not be performed again.

Tanner Stages will be summarized on a categorical scale by Visit. Shift table will be created to display shifts from baseline.

9.3.6 Menstruation

Caregivers will be asked if the female patient is menstruating and details will be recorded at Visit 1; any changes in normal cycles will be captured at Visit 8, Visit 10, and Visit 14 (end of treatment).

Menstruation details will be summarized as appropriate by visit.

9.4 Other Measures

9.4.1 Concomitant Medication

Medications will be coded using the World Health Organization Drug Dictionary WHO-DD Enhanced (v.WhoDrugGlobalB3 - 201903) or later version.

All medication will be considered concomitant.

For summaries and listings of medications the following approach will be used to determine the Anatomical Therapeutic Chemical (ATC) term to be presented:

- If coded to level 4 then the level 4 coded term will be presented.
- If coding is not performed at level 4 but level 3 coding is present then level 3 coded term will be presented.
- If coding is not performed at level 3 but level 2 coding is present then the level 2 coded term will be presented.
- If coding is not performed at level 2 but level 1 coding is present then the level 1 coded term will be presented.

Concomitant medications by ATC term and preferred term will be summarized by absolute counts (n) and percentages (%).

The ATC term, preferred term, reported generic name and reported brand name will be listed.

9.4.2 Procedures and Non-Drug Therapies

Procedures and non-drug therapies will be summarized overall.

9.4.3 Pregnancy Monitoring Form

Any patient who has become pregnant whilst receiving IMP, or within 90 days of last dose of IMP, must be reported to the GW PVD. Where possible the investigator should provide the outcome of the pregnancy and complete the pregnancy monitoring form. Details of pregnancies will be listed.

9.4.4 Patients Impacted by COVID-19

A listing of all patients impacted by COVID-19 related study disruption will be provided. The listing will include patient ID, site and a description of how the patient's participation was altered.

Protocol Deviations related to COVID will be provided in a data listing.

9.5 Changes in the Conduct of the Trial or Planned Analysis

After carefully evaluating the study performance and the potential impact of the COVID-19 pandemic to patients and site personnel the Sponsor made the decision to discontinue the lead-in trial. The significant challenges to conducting studies in this vulnerable group of patients were magnified by the COVID-19 pandemic leading to the conclusion that it was no longer feasible to conduct the studies. All patients enrolled in GWND18064 continued the study as per plan and if agreed, entered the OLE extension. When entering the OLE extension, patients will complete this trial as per protocol and annex. The annex described conditions for patients entering the extension trial if they withdrew from RCT trial due to Covid, due to sponsor administrative decision or delayed entry into GWND19002 due to Covid. All OLE trial Visit 1 data will be summarized for patients enrolling in the OLE under this protocol annex. Their data will be summarized in the same way as that of patients enrolling in the OLE under the GWND19002 protocol. The summary tables will present data from patients enrolling in the OLE under the protocol annex combined with that of patients enrolling in the OLE under the GWND19002 protocol. Their data will also be presented separately.'

As a result the analyses described in this SAP differ from those that were initially intended and presented in the protocol. Initially only comparison towards RCT baseline was planned, in this SAP both comparison with RCT and OLE baseline is planned.

10. References

1. US Federal Register. (1998) International Conference on Harmonization; Guidance on Statistical Principles for Clinical Trials. Department of Health and Human Services: Food and Drug Administration [Docket No. 97D-0174]. Federal Register Volume 63, Number 179, pages 49583-49598. September 16, 1998.

2. ASA. (2016) Ethical Guidelines for Statistical Practice. Prepared by the Committee on Professional Ethics, April 2016. http://www.amstat.org/about/ethicalguidelines.cfm

3. RSS. (2014) The Royal Statistical Society: Code of Conduct, 2014. http://www.rss.org.uk/Images/PDF/join-us/RSS-Code-of-Conduct-2014.pdf.

4. HealthActCHQ. (2013). The CHQ Scoring and Interpretation Manual. Boston, MA: HealthActCHQ.
5. Maruish, M. (2011). User's manual for the SF-36v2 Health Survey (3rd ed.). Lincoln, RI: QualityMetric Incorporated.

11. Tables, Listings, and Figures

Table, listings and figures shells are provided as a separate document. The final statistical tables will be produced in the format of the shells and will additionally include "double" page numbering in the format "page xx of yy". Note that programming notes may be added or modified if appropriate after each TLF shell.

The final statistical output will be provided as fully bookmarked pdf file including a table of contents.

All listings, tables, and graphs will have a header showing the sponsor company name and protocol and a footer showing the version of SAS, the file name and path, and the source of the data (eCRF page or listing number).

In general, one listing will be produced per eCRF domain. All listings will be sorted by site, and patient number.

In all listings a blank line will be placed between each patient. Within a data listing, if an item appears line after line (eg, repetition of patient number), then only the first occurrence will be displayed.

In data listings, the information for one patient will be kept on one page if at all possible, rather than splitting a patient's information across pages.

APPENDIX 1: Abbreviations

Abbreviation	Definition
AE	Adverse event
CRO	Contract research organization
CSR	Clinical study report
D	Day
eCRF	Electronical case report form
ITT	Intent-to –treat analysis set
ECG	Electrocardiogram
EMA	European Medicines aAgency
FDA	Food and Drug Administration
ICH	International Conference on Harmonisation
MedDRA	Medical ictionary for regulatory activities
OLE	Open-Label Extension
PD	Pharmacodynamics
PP	Per-protocol population
PK	Pharmacokinetics
QTc	QT-interval for ECG corrected for heart rate
SAE	Serious adverse event
SAP	Statistical analysis plan
SD	Standard deviation

Abbreviation	Definition
SAF	Safety population
SOC	System organ class
TEAE	Treatment-emergent adverse event
WHO	World Health Organization

APPENDIX 2 Adverse Events of Special Interest – Abuse Liability based on preferred term

Withdrawal	Drug withdrawal convulsions Drug withdrawal headache Drug withdrawal maintenance therapy Drug withdrawal syndrome Drug withdrawal syndrome neonatal Drug rehabilitation Rebound effect Steroid withdrawal syndrome Withdrawal arrhythmia Withdrawal syndrome
Drug abuse and dependence	Dopamine dysregulation syndrome Drug abuse Drug abuser Drug dependence Drug dependence, antepartum Drug dependence, postpartum Intentional drug misuse Intentional overdose Maternal use of illicit drugs Neonatal complications of substance abuse Polysubstance dependence Substance abuse Substance abuser Accidental overdose Dependence Disturbance in social behaviour Drug administered at inappropriate site Drug detoxification Drug diversion Drug level above therapeutic Drug level increased Drug screen Drug screen positive Drug tolerance Drug tolerance decreased Drug tolerance increased Medication overuse headache Narcotic bowel syndrome Needle track marks Overdose Prescribed overdose Prescription form tampering Substance use Substance-induced mood disorder Substance-induced psychotic disorder Toxicity to various agents

APPENDIX 3 Ranges for Clinically Significant Changes and Other Defined Flagged Values in Vital Signs

The range of values that will be used to identify clinically significant changes in vital signs parameters (See Section 9.3.1) are presented in Table 7.

Table 7 Ranges for Potentially Clinically Significant Changes in Vital Signs

Vital Sign	Range
Sitting Systolic BP (mmHg)	Change: < -20, > 20
Sitting Diastolic BP (mmHg)	Change: < -10, > 10
Pulse Rate (bpm)	Change: < -10, > 10
Weight (kg)	Percent Change: ≤ -7, ≥ 7

Defined flagged values that will be used to identify low or high vital signs parameters (See Section 9.3.1) are presented in Table 8.

Table 8 Other Defined Flagged Values for Vital Signs

Vital Sign	Flag
Sitting Systolic BP (mmHg)	< 90, > 140, > 160
Sitting Diastolic BP (mmHg)	< 50, > 90, > 100
Pulse Rate (bpm)	< 60, > 100
Temperature (°C)	> 38.0, < 36.0
Respiratory Rate (breaths/min)	< 12, > 20

APPENDIX 1 Defined Flagged Values in ECG Parameters

Defined flagged values that will be used to identify low or high ECG parameters (See Section 9.3.2) are presented in Table 9.

Table 9 Defined Flagged Values for ECG Parameters

ECG Parameter	Flag
QTc (ms)	> 450, > 480, > 500

APPENDIX 2 Toxicity Criteria for Laboratory Parameters

The toxicity criteria that will be used to identify abnormal laboratory parameters are presented in Table 10 and Table 11.

Table 10 Toxicity Criteria for Biochemistry Parameters

Parameter	Toxicity Decrease	Toxicity Increase
Chloride	≤0.96xLL	≥1.04xUL
Calcium	≤0.89xLL	≥1.16xUL
Sodium	≤0.96xLL	≥1.04xUL
Potassium	≤0.90xLL	≥1.10xUL
Glucose (mmol/L)	≤3.2	≥16
Phosphate	≤0.79xLL	
Cholesterol	≤0.85xLL	≥1.6xUL
AST		≥3xUL
ALT		≥3xUL
Lactate Dehydrogenase		≥2.6xUL
Alkaline phosphatase		≥2xUL
Gamma GT		≥2.6xUL
Bilirubin		>2xUL
Albumin	≤0.84xLL	
Total protein	≤0.84xLL	≥1.16xUL
Urea		≥2.6xUL
Blood urea nitrogen		≥2.6xUL
Creatinine		≥2.6xUL
Uric acid		≥1.16xUL

UL = upper limit of reference range LL = lower limit of reference range

Table 11 Toxicity Criteria for Hematology Parameters

Parameter	Toxicity Decrease	Toxicity Increase
Hemoglobin (g/dL)	≤9.4	
Hematocrit (%)	≤28	
Red cell count	≤0.84xLL	
Mean corpuscular volume	≤0.84xLL	≥1.11xUL
Mean corpuscular hemoglobin	≤0.84xLL	
Mean corpuscular hemoglobin concentration	≤0.84xLL	
Platelets (x10^9/L)	≤74	
Prothrombin time		>1.5xUL
Prothrombin international normalized ratio		>1.5
Total white blood cell count (x10^9/L)	≤2.9	≥21
Total neutrophil count (x10^9/L)	≤1.36	≥14.7
Segmented neutrophil count (x10^9/L)	≤0.75	≥12.3
Eosinophils (x10^9/L)		≥1.5
Basophils (x10^9/L)		≥0.31

Parameter	Toxicity Decrease	Toxicity Increase
Monocytes (x10^9/L)		≥2.1
Lymphocytes (x10^9/L) for patients <18 years (auto hematology)	≤1.0	
Lymphocytes (x10^9/L) for patients <18 years (manual hematology)	≤0.2	
Lymphocytes (x10^9/L) for patients ≥18 years	≤0.2	

UL = upper limit of reference range LL = lower limit of reference range

APPENDIX 5 Scoring the CHQ-PF50

The CHQ-PF50 provides information on various health concepts, as well as the PhS and PsS scales. Details for scoring each of these concepts and scales are based on the CHQ scoring and interpretation manual (HealthActCHQ, 2013).

Global Health Item

The global health item (item 1.1) of the CHQ-PF50 is calculated using the steps below.

1. The scores for this item are recoded as presented below, where missing scores remain missing.

Table 12 Scoring of Global Health Item

Response	Precoded Item Values	Final Item Values
Excellent	1	5
Very Good	2	4.4
Good	3	3.4
Fair	4	2.2
Poor	5	1

2. The final scores are then transformed so that they range from 0 to 100 in the following way:

$$(\text{Final item value} - 1) / 4 \times 100$$

Physical Function

Items 2.1 (a to f) are included in this health concept. Each item is coded with a value from 1 to 4, where 1 corresponds to a response "Yes, limited a lot", 2 "Yes, limited some", 3 "Yes, limited a little" and d 4 corresponds to the response "No, not limited".

The score for the physical function health concept is calculated using the following steps:

1. The mean of the scores (raw score) for items 2.1 (a to f) is calculated if at least 3 of the items have been answered. If less than 3 items have responses this concept is set to missing.
2. The standardized score which ranges from 0 to 100 is then calculated as follows:

$$(\text{Raw score} - 1) / 3 \times 100$$

Role/Social Limitations Due to Emotional or Behavioral Difficulties

This health concept is based on items 3.1 (a to c). Each item is coded with a value from 1 to 4, where 1 corresponds to a response "Yes, limited a lot" and 4 corresponds to the response "No, not limited".

The steps below are followed to score this health concept

1. If at least 2 of the 3 items have been answered, the mean of the scores (the raw score) for items 3.1 (a to c) is calculated, otherwise this concept is set to missing.
2. The standardized score which ranges from 0 to 100 is then calculated as follows:

$$(\text{Raw score} - 1) / 3 \times 100$$

Role/Social Limitations Due to Physical Health

Items 3.2 (a and b) are included in this health concept. Each item is coded with a value from 1 to

4, where 1 corresponds to a response "Yes, limited a lot" and 4 corresponds to the response "No, not limited".

The steps below are followed to score this health concept

1. The raw score is calculated as the mean of the scoresif at least 1 of the items has been answered, otherwise this concept is set to missing.
2. The standardized score which ranges from 0 to 100 is then calculated as follows:

(Raw score – 1) / 3 × 100

Bodily Pain and Discomfort

The Bodily Pain and Discomfort scale includes items 4.1 and 4.2. The scores for these items are to be reversed as presented in Table 13 and Table 14.

Table 13 Scoring of Item 4.1 for Bodily Pain and Discomfort

Item 4.1 Response	Precoded Item Values	Final Item Values
None	1	6
Very mild	2	5
Mild	3	4
Moderate	4	3
Severe	5	2
Very severe	6	1

Table 14 Scoring of Item 4.2 for Bodily Pain and Discomfort

Item 4.2 Response	Precoded Item Values	Final Item Values
None of the time	1	6
Once or twice	2	5
A few times	3	4
Fairly often	4	3
Very often	5	2
Every/almost every day	6	1

The Bodily Pain and Discomfort scale score is then calculated as follows:

1. The mean of the two scores (the raw score) is calculated if at least one of the items has been answered.
2. The standardized score which ranges from 0 to 100 is then calculated as follows:

(Raw score – 1) / 5 × 100

Behavior

Items 5.1 (a to e) and 5.2 are included in this health concept. Items 5.1 (a to e) are coded with a value from 1 to 5, where 1 corresponds to a response "Very often", 2 to "Fairly often", 3 to "Sometimes", 4 to "Almost never", and 5 corresponds to the response "Never".

Responses to item 5.2 are recoded using the rules in Table 15.

Table 15 Scoring of Behavior

Response	Precoded Item Values	Final Item Values
Excellent	1	5
Very Good	2	4.4
Good	3	3.4
Fair	4	2.2
Poor	5	1

The following steps are then used to calculate the Behavior concept score:

1. If at least 3 items have been answered the mean is calculated, otherwise this scale is set to missing.
2. The resulting mean, or raw score, is then standardized to range from 0 to 100 as follows:

$$(\text{Raw score} - 1) / 4 \times 100$$

Global Behavior Item

The global behavior item (item 5.2) of the CHQ-PF50 is calculated using the steps below.

1. The scores for this item are recoded as presented below, where missing scores remain missing.

Table 16 Scoring of Global Behavior Item

Response	Precoded Item Values	Final Item Values
Excellent	1	5
Very Good	2	4.4
Good	3	3.4
Fair	4	2.2
Poor	5	1

2. The final score is then transformed to range from 0 to 100 in the following way:

$$(\text{Final item value} - 1) / 4 \times 100$$

Mental Health

Items 6.1 (a to e) are included in this health concept. Each of the items 6.1 (a to d) is coded with a value from 1 to 5, where 1 corresponds to a response of "All of the time" and 5 corresponds to the response "None of the time". Responses to item 6.1 (e) are reverse scored as indicated in Table 17.

Table 17 Scoring of Mental Health

Response	Precoded Item Values	Final Item Values
All of the time	1	5
Most of the time	2	4
Some of the time	3	3
A little of the time	4	2
None of the time	5	1

If at least 3 of the items 6.1 (a to e) have been answered the raw score is obtained by calculating the mean of the scores. The raw score is then standardized to range from 0 to 100 as follows:

(Raw score – 1) / 4 × 100

Self Esteem

The self esteem scale includes Items 7.1 (a to f) which are reverse scored as presented in Table 18.

Table 18 Scoring of Self Esteem

Response	Precoded Item Values	Final Item Values
Very satisfied	1	5
Somewhat satisfied	2	4
Neither satisfied nor dissatisfied	3	3
Somewhat dissatisfied	4	2
Very dissatisfied	5	1

If at least 3 of the items 7.1 (a to f) have been answered the mean of the scores is calculated. The mean score is then standardized to range from 0 to 100 as follows:

(Raw score – 1) / 4 × 100

General Health Perceptions

Items 8.1 (a to e) and item 1.1 are included in this scale. Items 8.1 (a, c, e) are scored from 1 to 5 where 1 represents the response "Definitely true" and 5 represents "Definitely False". Items 8.1 (b and d) are reverse scored as is seen in Table 19.

Table 19 Scoring of Global Health Perceptions

Response	Precoded Item Values	Final Item Values
Definitely true	1	5
Mostly true	2	4
Don't know	3	3
Mostly false	4	2
Definitely false	5	1

Finally, responses to item 1.1 are recoded as presented in Table 13.

The general health perceptions scale is then scored using the steps below:

1. If less than 3 of the items 8.1 (a to e) and 1.1 have been answered, this scale is set to missing. Otherwise calculate the raw score by computing the mean of the responses.
2. The raw score is then standardized so that it ranges from 0 to 100 as follows:

(Raw score – 1) / 4 × 100

Change in Health Item

The change in health item (item 8.2) of the CHQ-PF50 is to be reverse scored as displayed in Table 20.

Table 20 Scoring of Change in Health Item

Response	Precoded Item Values	Final Item Values
Much better now than 1 year ago	1	5
Somewhat better now than 1 year ago	2	4
About the same now than 1 year ago	3	3
Somewhat worse now than 1 year ago	4	2
Much worse now than 1 year ago	5	1

This is a categorical scale and thus no transformation is necessary.

Emotional Impact on Parent

This scale includes items 9.1 (a to c) which are to be reverse scored in accordance with Table 21.

Table 21 Scoring of Emotional Impact on Parent

Response	Precoded Item Values	Final Item Values
Not at all	1	5
A little bit	2	4
Some	3	3
Quite a bit	4	2
A lot	5	1

The emotional impact on parent scale score is then calculated as follows:

1. If at least 2 of the items in this scale have been answered, the raw score is obtained by computing the mean of the responses, otherwise this scale is set to missing.
2. The raw score is then standardized so that it ranges from 0 to 100 as follows:

$$(\text{Raw score} - 1) / 4 \times 100$$

Time Impact on Parent

Items 9.2 (a to c) are included in the time impact on parent scale. The responses range from 1 to 4 where 1 corresponds to "Yes, limited a lot" and 4 corresponds to a response of "No, not limited". If at least 2 items in this scale have been answered the raw score is calculated as the mean of the items. If less than 2 items have been answered the scale is set to missing.

Finally, the standardized score which ranges from 0 to 100 is calculated as

$$(\text{Raw score} - 1) / 3 \times 100$$

Family Activities

The family activities health concept is comprised of items 9.3 (a to f). The responses range from 1 to 5 where 1 represents the response "Very often" and 5 the response of "Never". If at least 3 items in this scale the raw score is calculated as the mean of the responses, otherwise the scale is sent to missing.

The raw score is then standardized as:

$$(\text{Raw score} - 1) / 4 \times 100$$

Family Cohesion Item

The family cohesion health concept includes item 9.4 which is recoded using the rules outlined inTable 22.

Table 22 Scoring of Family Cohesion Item

Response	Precoded Item Values	Final Item Values
Excellent	1	5
Very good	2	4.4
Good	3	3.4
Fair	4	2.2
Poor	5	1

The score is then transformed so that it ranges from 0 to 100 as follows:

$$(\text{Raw score} - 1) / 4 \times 100$$

Standardized Physical Summary and Standardized Psychosocial Summary

The PhS and PsS scales are scored using norm-based methods based on data from from the general U.S. population and six clinical samples of children. Table 24 includes the sample means and standard deviations, as well as factor score coefficients, used to derive the PhS and PsS scale scores.

Table 24 Information Required for Scoring PhS and PsS Scales

CHQ-PF50 SCALE*	MEAN	SD	FACTOR SCORE COEFFICIENTS	
			PhS	PsS
PF	90.8525408	16.3826344	.37138	-.09243
RP	91.4951246	18.9079749	.34493	-.06973
GH	66.6958379	19.3564297	.29460	-.05547
BP	78.6833515	20.7355708	.27883	-.05514
REB	90.4013015	19.5067502	-.01178	.21155
PT	83.8816188	20.2901603	.09113	.16944
PE	73.9788476	21.406013	.06063	.19823
SE	79.2555314	17.8308361	-.09480	.24792
MH	77.2595806	13.6861999	-.08263	.25335
BE	72.3086051	17.1447913	-.12675	.27911

Note: PF = Physical Functioning; RP = Role/Social-Physical; GH = General Health Perceptions; BP = Bodily Pain; REB = Role/Social Emotional/Behavioral; PT = Parental Impact-Time; PE = Parental Impact-Emotional; SE = Self Esteem; MH = Mental Health; BE = Behavior

The steps used in calculating PsS and PhS are summarized below:

1. Calculate each of the CHQ-PF50 scales presented in Table and using the methods outlined in this appendix.
2. Each scale is then standardized using a z-score transformation as follows:

$$\frac{scale\ score - mean\ of\ scale}{std\ deviation\ of\ scale}$$

   where the mean and standard deviation for each scale are obtained from Table 24
3. Calculate PhS and PsS as a weighted sum of each standardized scale, using the factor score coefficients from Table 24 as the weights for each item in each scale.
4. Transform each scale score to obtain t-score with a mean of 50 and standard deviation of 10 as follows:

   PhS = PhS weighted sum × 10 + 50

   PhS = PhS weighted sum × 10 + 50.