Statistical Analysis Plan

Sponsor	GW Research Ltd
Protocol Title:	A RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED TRIAL TO INVESTIGATE THE EFFICACY AND SAFETY OF CANNABIDIOL ORAL SOLUTION (GWP42003-P, CBD-OS) IN PATIENTS WITH RETT SYNDROME
Protocol Number:	GWND18064
Premier Research PCN:	GWRE166121
EudraCT number:	2018-003370-27
ClinicalTrials.gov identifier	NCT03848832
Document Version:	Final 2.0
Document Date:	30-MAR-2021

Approvals

Role	Signatures	Date(dd-Mmm-yyyy)
Biostatistician	Print Name: [redacted] Senior Manager, Biostatistics, Premier Research	30-Mar-2021 | 09:21:47 EDT
	Sign Name: DocuSigned by: [redacted] Signer Name: [redacted] Signing Reason: I approve this document Signing Time: 30-Mar-2021 | 09:21:41 EDT 452890CD70774C35B1D116346F9E649B	
GW Research Ltd Representative	Print Name: [redacted] Senior Manager, Biostatistics	31-Mar-2021 | 00:19:21 PDT
	Sign Name: DocuSigned by: [redacted] Signer Name: [redacted] Signing Reason: I approve this document Signing Time: 30-Mar-2021 | 06:27:58 PDT B728BF44216240D482871298B007CB1B	

Document History

Version	Date	Author	Description
Final 1.0	02FEB2021	[redacted]	Initial Version
Final 2.0	30MAR2021	[redacted]	Update to add clarification on: - Derivation of treatment received - Imputation of AE dates for time to onset/resolution calculation - Exposure outputs to produce - AE analysis based on resolution - Classification of ALT/AST - ECG triplicates analysis

Table of Contents

List of Tables

1. Overview

This statistical analysis plan (SAP) describes the planned analysis and reporting for GW protocol number GWND18064 (A randomized, double-blind, placebo-controlled trial to investigate the efficacy and safety of cannabidiol oral solution (GWP42003-P; CBD-OS) in patients with Rett syndrome (RTT), dated 07 Jul 2020 version #4.1. Reference materials for this statistical plan include the protocol and the accompanying sample data collection documents. 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. The concept of statistical analysis has been changed, no inferential analysis will be performed in the study, the statistical analysis will be purely descriptive. The reason for this, as described in Section 8.5, is that due to potential impact of the COVID-19 pandemic to patients and site personnel the Sponsor made the decision to discontinue the trial.

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 unblinded inferential or descriptive analysis of data pertaining to study GWND18064.

2. Study Objectives and Endpoints

2.1 Study Objectives

2.1.1 Primary Objective

The primary objective is to evaluate the efficacy of 15 mg/kg/day GWP42003-P, compared with placebo, at the end of 24 weeks' treatment in reducing symptom severity in patients with RTT using the Rett Syndrome Behaviour Questionnaire (RSBQ).

2.1.2 Secondary Objectives

Key Secondary Objective:

- To evaluate the efficacy of 15 mg/kg/day GWP42003-P, compared with placebo, at the end of 24 weeks' treatment in reducing symptom severity in patients with RTT using the Clinical Global Impressions - Improvement (CGI-I).

Other Secondary Objectives:

- To evaluate the efficacy of 5 mg/kg/day GWP42003-P, compared with placebo in:
  - RSBQ
  - CGI-I
- To evaluate the effect of GWP42003-P, compared with placebo, in other measures of disease severity.
  - RSBQ subscales.
  - Clinical Global Impressions - Severity (CGI-S).
  - 9-items Motor Behavioral Assessment (MBA-9).
  - Children's Sleep Habits Questionnaire (CSHQ).
- To evaluate the safety of GWP42003-P, compared with placebo, in patients with RTT.

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).
- To determine the plasma concentrations of cannabidiol (CBD) and its major metabolites with the aim of evaluating exposure versus efficacy and safety and collecting data for population pharmacokinetic (PK) analysis.
- To evaluate the effect of GWP42003-P on exploratory biomarkers.

2.2 Study Endpoints

2.2.1 Efficacy Endpoints

2.2.1.1 Primary Efficacy Endpoint

The primary efficacy endpoint is the RSBQ (total score) for the 15 mg/kg/day GWP42003-P dose level compared with placebo at the end of 24 weeks.

2.2.1.2 Secondary Efficacy Endpoints

Key Secondary Endpoint:

- CGI-I for the 15 mg/kg/day GWP42003-P dose level compared with placebo at the end of 24 weeks.

Other Secondary Efficacy Endpoints:

- To compare 5 mg/kg/day GWP42003-P with placebo using:
  - RSBQ (total score)
  - CGI-I

- To compare 15 mg/kg/day GWP42003-P and 5 mg/kg/day GWP42003-P with placebo using:
    - RSBQ subscales.
    - CGI-S.
    - MBA-9.
    - CSHQ.

2.2.1.3 Exploratory Efficacy Endpoint(s)

- To compare 15 mg/kg/day GWP42003-P and 5 mg/kg/day GWP42003-P with placebo for the following endpoints:
    - Caregiver QoL questionnaire (SF-36).
    - Patient QoL questionnaire (CHQ-PF50).
    - Hospital Services Use Questionnaire.
    - Caregiver assessment of Rett symptoms (symptom diary).
    - Blood levels of exploratory biomarkers.

2.2.1.4 Other Exploratory Endpoint(s)

- Plasma concentrations of CBD and its main metabolites.

2.2.2 Safety Endpoints

The safety profile of GWP42003-P compared with placebo will be assessed by measuring:

- 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).

3. Overall Study Design and Plan

3.1 Overall Design

This multisite trial consists of a double-blind, randomized, placebo-controlled design which will compare the efficacy of GWP42003-P versus placebo over a 24-week treatment period.
Patients will be randomized to receive 5 mg/kg/day GWP42003-P, 15 mg/kg/day GWP42003-P or matching volumes of placebo in a 2:2:1:1 ratio.
Following screening, eligible patients will complete a 2 to 4-week baseline period (patients can be randomized as soon as they have completed at least 2 weeks of baseline and all eligibility criteria are confirmed [including review of all clinical laboratory results and, if applicable, confirmation of *MECP2* pathogenic genetic mutation]). All patients entering the trial will be stratified by severity, based on their Clinical Severity Scale (CSS) score (either a CSS score within the range of 10–22, or a CSS score within the range 23–36). Patients will be randomized and will commence the 24-week treatment period, including up to 2- weeks dose escalation. The treatment period ends at Visit 9 (Day 169), after which patients will commence a 10-day taper period followed by the 4-week follow-up period. The 10-day taper period and 4-week follow-up period may not be required for patients continuing CBD-OS treatment under a separate protocol.
If a patient permanently discontinues treatment at any point during the trial, the investigational medicinal product (IMP) should be gradually reduced over 10 days (unless inadvisable due to an AE). Patients and caregivers will be encouraged to remain in the trial and continue to complete trial assessments and visits as per protocol.
If a patient withdraws from the trial at any point during the trial, they will be required to attend a withdrawal visit, if applicable taper the IMP and attend the end of taper visit, and then complete the 4-week follow-up period.

3.2 Sample Size and Power

252 patients will be randomized to receive 1 of 2 dose levels of GWP42003-P (5 mg/kg/day or 15 mg/kg/day) or matching placebo (5 mg/kg/day dosing volumes or 15 mg/kg/day dosing volumes) on a 2:2:1:1 basis. The placebo data for the 2 dose groups will be pooled for analysis. Pooling of the placebo dose groups will be implemented for efficacy and safety analyses. For safety analysis, summaries by placebo dose will also be presented.

Assuming a common standard deviation of 9.38 and using a 2-sided test at a 0.05 α-level, a total sample size of 252 patients (84 patients per active dose group and 42 patients per placebo group) will provide 90% power to detect a mean difference of 5 points in change from baseline to Visit 9 in RSBQ (total score) between 15 mg/kg/day of GWP42003-P and placebo, allowing for 10% withdrawals.

3.3 Study Population

Female or male patients aged 2–18 years (inclusive), who weigh at least 10 kg may enter the 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 complete list of inclusion/exclusion criteria, please consult clinical study protocol.

3.4 Treatments Administered

The administered treatment is GWP42003-P oral solution or placebo 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 in the clinic at the end of Visit 2 (Day 1) and caregivers will be instructed how to measure and administer the IMP to the patient.

Patients will escalate IMP to the target dose level for their dose group over a period of up to 2 weeks depending on the target dose (e.g., no titration is required to 5 mg/kg/day while titration to 15 mg/kg/day takes 2 weeks [weekly increments of 5 mg/kg/day]). Following escalation patients will remain at their assigned target dose level for the duration of the maintenance period of the trial.

3.5 Method of Assigning Subjects to Treatment Groups

At the start of Visit 1 (screening), enrolled patients will be allocated a unique patient number. After completion of assessments and confirmation of eligibility at Visit 2, patients will be randomized to receive 5 mg/kg/day GWP42003-P, 15 mg/kg/day GWP42003-P or matching volumes of placebo in a 2:2:1:1 ratio. All patients entering the trial will be stratified by severity, based on their Clinical Severity Scale (CSS) score (either a CSS score within the range of 10–22, or a CSS score within the range 23–36). GW will provide all IMP in a packed and labeled state and the RTSM system will identify the pack number(s) to be dispensed to the patient at each relevant visit, according to the treatment assigned in the randomization schedule.

3.6 Blinding and Unblinding

The identity of IMP assigned to patients will be held by the RTSM system. The principal investigator (PI) at each site, or his/her designee, is responsible for ensuring that information on how to access the RTSM system is available to the relevant staff in case of an emergency and unblinding is required. A patient's treatment assignment should only be unblinded when knowledge of the treatment is essential to make a decision on the medical management of the patient. Unblinding for any reason will be considered a protocol deviation.

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**	**3[a]**	**4[a]**	**5**	**6**	**7**	**8**	**9[b,c]**	**10[d]**	**11[a]**
Day Number (Visit window)	**−28 to −14***	**1 (+3)**	**8 (± 3)**	**15 (± 3)**	**29 (± 3)**	**57 (± 3)**	**85 (± 3)**	**127 (± 3)**	**169 (± 3)**	**179 (+3)**	**207 (+3)**
Informed consent and assent[e]	X										
Demographics	X										
Medical history	X										
CSS	X										
Eligibility check	X	X									
Genetic analysis of *MECP2* (if unknown)[f]	X										
Randomization		X									
Concomitant medications	X	X	X	X	X	X	X	X	X	X	X
Adverse events	X	X	X	X	X	X	X	X	X	X	X
Menstruation question (where appropriate)		X							X		
Physical examination (including weight)	X				X		X		X		
Height	X								X		
ECG	X				X	X	X	X	X		
Vital signs	X	X			X	X	X	X	X	X	
Clinical laboratory blood sampling (hematology and biochemistry)[g]	X				X	X	X	X[h]	X		
Dipstick urinalysis (where possible)	X								X		
Visit Number	**1**	**2**	**3[a]**	**4[a]**	**5**	**6**	**7**	**8**	**9[b,c]**	**10[d]**	**11[a]**
Day Number (Visit window)	**−28 to −14***	**1 (+3)**	**8 (± 3)**	**15 (± 3)**	**29 (± 3)**	**57 (± 3)**	**85 (± 3)**	**127 (± 3)**	**169 (± 3)**	**179 (+3)**	**207 (+3)**
Urine/Serum THC screen[i]	X										
Serum pregnancy test (if appropriate)	X								X		
PK blood sampling (IMP)					X[j,k,l]	X[j,k,l]	X[j,k,l]		X[j,k,l]		

Biomarker blood sampling		X								X		
Caregiver completed questionnaire	RSBQ	X	X			X	X	X	X	X		
	CSHQ		X			X	X	X	X	X		
	Caregiver QoL questionnaire (SF-36)		X							X		
	Patient QoL questionnaire (CHQ-PF50)		X							X		
	Symptom diary[n]	X	X	X	X	X	X	X	X	X		
Time and date of last IMP dose						X	X	X		X		
MBA-9			X					X		X		
CGI-S			X			X	X	X	X	X		
CGI-I						X	X	X	X	X		

Visit Number	**1**	**2**	**3[a]**	**4[a]**	**5**	**6**	**7**	**8**	**9[b,c]**	**10[d]**	**11[a]**
Day Number (Visit window)	**−28 to −14***	**1 (+6)**	**8 (± 3)**	**15 (± 3)**	**29 (± 3)**	**57 (± 3)**	**85 (± 3)**	**127 (± 3)**	**169 (± 3)**	**179 (+3)**	**207 (+3)**
Hospital Services Use Questionnaire		X			X	X	X	X	X		
Tanner Staging (where appropriate)	X								X		
Suicidality assessment	X				X	X	X	X	X	X	
IMP dispensing[o]		X			X	X	X	X	X		
IMP collection and compliance review					X	X	X	X	X	X	
Dosing diary[p]		X	X	X	X	X	X	X	X	X	

*Patients can be randomized as soon as they have completed at least 2 weeks of baseline and all eligibility criteria are confirmed (review of clinical laboratory results and, if applicable, genetic analysis)

a Visit to be conducted by telephone.
b To be performed for all patients completing or withdrawing from the trial. Patients who withdraw early should commence the 10-day IMP taper period, if possible.
c A safety follow-up visit 4 weeks after last IMP dose is required for all patients who withdraw from the trial or complete the trial but do not enroll in the OLE.
d Only required for those patients who do not participate in the OLE trial or for those who withdraw from the trial early and taper IMP. For patients

who complete treatment but do not participate in the OLE trial, Visit 10 should be 10 (+3) days after Visit 9. For patients who withdraw early and taper IMP, this visit should occur 10 (+3) days after the withdrawal visit.

e 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 parental/legal representative consent.

f For patients weighing < 13 kg and without documented pathogenic *MECP2* mutation, a blood sample for pathogenic *MECP2* genetic mutation genetic analysis must be in advance of Visit 1 (following informed consent [and where applicable, patient assent]).

g Determination of serum IGF-1 levels at Visit 1 and 9 only. IGF-1 laboratory results will remain blinded throughout the trial.

h For patients taking concomitant valproic acid only.

i THC screen using urine or serum , may be carried out at the site using a drug of abuse urine dipstick; otherwise a urine or blood sample at Visit 1 may be analyzed and sent to the central laboratory (note: patient weighing < 12 kg, must not provide a serum sample can be collectedfor THC screen at Visit 1 instead; if there are difficulties obtaining sufficient urine for THC testing atVisit 1 , then a blood or urine sample must be collected at the site on another day within ± 7 days of Visit 1).

j To be taken together with the safety laboratory sample collection, as appropriate considering the patient's weight (patient must weigh ≥ 12 kg).

k Trough sample at Visit 9 (and preferentially also trough sample at Visits 5, 6 and 7, if possible) i.e., collected 9–18 hours after IMP dose. Caregivers should be advised not to administer the IMP prior to attending Visit 9 (and preferentially advised not to administer IMP prior to attending Visits 5, 6, and 7, if possible). The IMP can then be administered during the visit, once the PK sample has been collected.

l The time of the patients' sample collection, time of latest IMP dose, time and type of meal consumed by the patient closest to the latest IMP dose will be recorded, as well as the time of latest concomitant medications.For patients weighing < 16 kg, no blood sample will be collected for biomarker analysis.

m Completed weekly throughout the trial.

n In cases where patients are not able to attend study visits due to special circumstances (e.g., COVID-19 pandemic), the investigator will discuss with the Sponsor potential mitigation approaches for IMP dispensing, secure delivery, and collection.

o Completed daily throughout the trial.

4. Statistical Analysis and Reporting

Statistical analyses will be performed after the database is locked. Only after the database has been locked will the unblinded statistician make the treatment allocations available.

4.1 Introduction

Data processing, tabulation of descriptive statistics, calculation of inferential 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 for the treatment groups, unless otherwise specified.

The number of missing values will be calculated as difference between the total number of patients in the study population for the treatment group 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 blinded data review meeting will take place. After database lock, the study will be unblinded and descriptive statistical outputs will be created using the unblinded data.

5. Analysis Populations

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

Screened

- All patients screened.

Randomized

- All patients randomized.

Intent-to treat set (ITT)

- All patients who are randomized, receive at least 1 dose of IMP in the trial, and have baseline efficacy data will be included and analyzed according to their randomized treatment arm.
- The ITT set is the primary analysis set for all efficacy endpoints.

Safety Analysis Set

- All patients who receive at least 1 dose of IMP in the trial will be included and analyzed according to the treatment received. Only patients for whom it has been confirmed that they did not take any IMP will be excluded from this safety analysis set. This analysis set will be used to report the safety data. Treatment received will be derived using the real kits dispensed (reported in database) as following:
  - If patient took only placebo, assign to placebo.
  - If patient took mostly placebo but one dose of 5 mg/kg/day (or 15 mg/kg/day) then assign to 5 mg/kg/day (or 15 mg/kg/day).
  - If patient was assigned to CBD and took CBD then assign them to the dose group they were meant to be assigned to.

Analysis sets will be identified prior to database lock and unblinding of the trial data.

6. General Issues for Statistical Analysis

All data collected in the study will be listed, ordered by site, treatment, 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 Study treatments

Table 2 Study treatments

Endpoint	Actual Treatment	Treatment Label (Table column/Figure legend)	Sorting order in listings, column order in tables
Efficacy	Pooled Placebo	Placebo	3
Safety, Demographics and Baseline	5 mg/kg/day Placebo	Placebo 5 mg/kg	3
	15 mg/kg/day Placebo	Placebo 15 mg/kg	4
	Pooled Placebo	Placebo	5
All	5 mg/kg/day GWP42003-P	GWP42003-P 5 mg/kg	1
All	15 mg/kg/day GWP42003-P	GWP42003-P 15 mg/kg	2

For safety tables where placebo is split by dosage, an additional Pooled Placebo column will be included.

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
Visit 1: Screening	Screening
Visit 2: Day 1, baseline visit	Day 1
Visit 3: Day 8	Day 8
Visit 4: Day 15	Day 15
Visit 5: Day 29	Day 29
Visit 6: Day 57	Day 57
Visit 7: Day 85	Day 85
Visit 8: Day 127	Day 127
Visit 9: Day 169	End of Treatment
Visit 10: Day 179	End of Taper
Visit 11: Day 207	Safety 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 efficacy endpoints Visit 2 will be used as the baseline observation for all calculations. The baseline value for the symptom diary will be the mean of the symptom diary scores prior to first dose.

For all safety endpoints, the last non-missing observation recorded prior to the first drug administration will be used as the baseline observation for all calculations.

6.1.3 Last Visit

The last visit for efficacy endpoints assessed at clinic visits is defined as the last scheduled visit at which a patient's last evaluation (up to Visit 9/end of treatment visit/any other last visit) is performed.

6.1.4 Study Periods

Treatment period: Day 1 to Day 169.

For 15 mg/kg/day group only:

Titration period: Day 1 to Day 14

Maintenance period: Day 15 to Day 169.

6.1.5 Day Numbering

The first day of treatment (Day 1) will be the date on the 'First IMP Dose' eCRF.

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)

6.1.6 Multiple Comparisons

Multiple comparisons with regards to primary/key secondary/secondary endpoints are detailed in Efficacy Analysis section 8.

6.1.7 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.

6.1.7.1 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.

6.1.7.1.1 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.

6.1.7.1.2 CHQ-PF50

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

6.1.7.1.3 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 2 and date of Visit 9, and events occurring on date of Visit 2 will be assumed post-IMP as per Section 6.1.9.1 Stop dates will not be imputed if the AE is ongoing.

The imputation method will be used to determine treatment emergence and time to onset/resolution of an AE, and imputed dates/times will not be presented in AE outputs.

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'.

6.1.7.1.4 Concomitant Medication

Missing concomitant medication dates will be handled in a similar fashion as described for AEs in Section 6.1.7.1.3.

6.1.8 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.

6.1.9 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.

6.1.9.1 Adverse events

A TEAE is defined as an AE with a start date on or after the first dose of IMP. If an AE has a partial start date and it is unclear from the partial date (or the stop date) whether the AE started prior to or post first dose of IMP then the AE will be considered treatment emergent. If the start date of the AE is the same as the date of first dose of IMP then the AE will be considered treatment emergent.

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'.

An AE 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 – Date of first dose of IMP + 1

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

Stop date of AE – Start date of AE + 1

6.1.9.2 Prior and concomitant medications

Medications that started and stopped prior to the first study drug intake will be considered prior medications. A concomitant medication is defined as any medication that was administered during the treatment period. This includes medications that started before the treatment period and continued while on treatment and medications that started during the treatment period.

If the start/stop dates of a medication are partially or completely missing, then the medication will be assumed to be concomitant if it cannot be definitely shown that it was not administered during the treatment period. Missing dates will not be replaced.

6.1.9.3 Age

Age will be calculated as:

(Date of screening – date of birth) ÷ 365.25.

6.1.9.4 Exposure

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

(Date of last dose in the treatment period – Date of Day 1) + 1

6.1.10 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 by treatment group and overall using standard summary statistics. The number screened, number of screen failures and number randomized will be included.

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

Patient disposition, including patients treated, patients who completed the treatment period and the taper phase, patients entering the OLE, patients discontinued (including reason for discontinuation) from the treatment but completed trial, patients that withdrew from trial, and taper phases will be summarized by absolute counts (n) and percentages (%). A further table split by site, and by country will be produced, showing number of patients randomized, withdrawn and completed the treatment period and entering the OLE 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 randomized 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 by treatment group and overall for the safety and ITT analysis sets:

- Age (years);
- Age group (2-5 years, 6-12 years and 13-18 years);
- Sex;
- Race;
- Ethnicity
- Country;
- Region (United States, Rest of the World);
- Weight at baseline (kg);
- Height at baseline (cm);
- Body mass index at baseline (kg/m2);

The following baseline characteristics will be summarized by treatment group and overall for the safety and ITT analysis sets:

- CSS Score;
- CSS stratum;
- Previous use of cannabis.

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.

The number of patients with relevant or significant medical or surgical history and medical 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 specific treatment group. Two tables will be produced, one including any events classified as resolved at screening, and the other including all current conditions.

7.4 Exposure and Compliance

IMP is to be administered twice daily (morning and evening). The first dose will be taken in the clinic on Day 1. The date of final dose in the treatment period will be recorded on the eCRF. The date of final dose, for patients who enter the taper period, will be recorded on the eCRF at the end of taper visit.

For dosing during the treatment period, the following data will be presented: total number of dosing days, number of days with diary reported, number of days with doses not taken according to the diary will be reported as continuous variables. Number and percentagesof dose adjustments (escalation, decrease, other) as well as methods of IMP administration will be described.

8. Efficacy Analysis

The primary analyses will use the ITT analysis set.

No formal hypothesis testing will be carried out for this study. Statistical analysis will be descriptive.

8.1 Primary Efficacy Endpoint

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 will be summarized on a numerical scale by visit and treatment group.

8.2 Key Secondary Efficacy Endpoint

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 and treatment group.

The analyses described above will be performed on the ITT analysis set.

8.3 Other Secondary Efficacy Endpoints

8.3.1 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 treatment and visit for each subscale.

The analysis will be performed on the ITT analysis set.

8.3.2 Caregiver Global Impression of Severity

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 will be summarized by treatment and visit. Frequency table will be created for CGIS categories by treatment and visit.

8.3.3 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 will be summarized by treatment and visit.

8.3.4 Children's Sleep Habit Questionnaire

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 5 CHSQ 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 will be summarized by treatment and visit.

8.4 Exploratory Efficacy Endpoints

The analysis of exploratory efficacy endpoints will be descriptive, using summaries by treatment arms.

8.4.1 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 6 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 will be summarized by treatment and visit.

8.4.2 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 7 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 will be summarized by treatment and visit.

8.4.3 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 summarized by treatment group for the baseline and treatment periods.

8.4.4 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 will be summarized by visit and treatment group.

9. Safety and Tolerability Analysis

9.1 Adverse Events

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

Summaries will be presented by treatment group as well as SOC and preferred term.

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

Overall summary of AEs, including number of patients reporting each of; Treatment Emergent Adverse Events (TEAEs), TEAEs that started during the titration period as applicable, treatment-related TEAEs, treatment-related TEAEs that started during the titration period, TEAEs leading to withdrawal, treatment-related TEAEs leading to withdrawal, serious TEAEs, treatment-related serious TEAEs.

- Summary of TEAEs.
- Summary of TEAEs by event (PT).
- Summary of treatment-related TEAEs.
- Summary of treatment-related TEAEs by event
- Summary of TEAEs by maximal severity.
- Summary of TEAEs by sex.
- Summary of serious TEAEs.
- Summary of serious TEAEs by event.
- Summary of non-serious TEAEs.
- Summary of non-serious TEAEs by event (PT).
- Summary of treatment-related serious TEAEs.
- Summary of treatment-related serious TEAEs by event (PT).
- 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 temporary IMP dose reduction (by resolution and overall).
- Summary of treatment-related TEAEs leading to temporary IMP dose reduction (by resolution and overall).
- Summary of TEAEs leading to permanent IMP dose reduction excluding permanent discontinuation (by resolution and overall).
- Summary of treatment-related TEAEs leading to permanent IMP dose reduction excluding permanent discontinuation (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) in the GWP42003-P treatment groups and where the incidence is greater than the (pooled) placebo treatment group.
- List of patients experiencing TEAEs by SOC and preferred term.
- Summary of pre-treatment AEs.

For the summary of TEAEs by maximal severity, for each patient, the worst severity recorded by preferred term, 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 'Continuing'. For each patient, preferred term and SOC, if both 'Resolved' and 'Continuing' AE are present in database, only the 'Continuing' one will be presented in outputs.

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

- Week 1 (Day 1–7).
- Week 2 (Day 8–14).
- Weeks 3–6 (Day 15–42).
- Weeks 7–10 (Day 43–70).
- Weeks 11–14 (Day 71–98).
- Weeks 15–18 (Day 99–126).
- Weeks 19–22 (Day127–154).
- >22 weeks (> Day 154).

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 the first dose of IMP (as recorded on the eCRF) will be calculated as per Section 6.1.56.1.5. For partial dates, if it is clear from the partial date that the start/stop day was prior to the first dose of IMP, then 'pre' will be listed, similarly if it is clear that the event was post the first dose of IMP then 'post' will be listed as the start/stop day as appropriate. If it is not clear whether it is pre or post first dose of IMP, then day is not calculated, field is left blank.

All AEs will be listed. Listings will include the start and stop day of the AE, a flag for treatment emergence, and limited demographic information about the patient (age, sex, race and weight at screening). A separate listing will be provided for pre-treatment AEs, serious AEs and events of special interest (see APPENDIX Appendix 2).

9.2 Clinical Laboratory Evaluation

9.2.1 Hematology and Biochemistry

Hematology and biochemistry safety parameters are measured at Visit 1 (screening), Visit 5, Visit 6, Visit 7, Visit 8 (for patients taking valproic acid only) and Visit 9 (end of treatment).

Summaries will be presented by treatment group for each laboratory parameter at each visit. Change from 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.

Shift tables for hematology and biochemistry parameters will be produced, based upon normal ranges and GW toxicity limits (See Appendix 4), to determine the categorical shifts from baseline 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. For each of these parameters, a patient with value > 8 ULN will also fall into the >3 and >5 ULN categories.

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 the same visits as biochemistry and hematology.

Urinalysis results will be listed only.

9.2.3 Pregnancy Test and Urine THC Screen

Serum pregnancy test results and urine THC screen results will be summarized by treatment group and visit.

9.2.4 MECP2

Results from the *MECP2* genetic mutation assessment at screening will be listed, along with demographic data in one listing. The listing will indicate whether the *MECP2* genetic mutation was known at screening or whether it was obtained from laboratory results.

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 (screening), Visit 2 (Day 1), Visit 5, Visit 6, Visit 7, Visit 8, Visit 9 (end of treatment) and Visit 10 (end of taper).

Summaries will be presented by treatment group for each vital sign parameter at each visit. Change from baseline to each post-baseline visit will also be presented.

A separate table will be produced, by treatment group and 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 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, visit and treatment group. The number of patients with at least one post-baseline flagged vital sign parameter value will be summarized by parameter, flagged criteria and treatment group.

9.3.2 Electrocardiogram

An ECG will be performed at Visit 1 (screening), Visit 5, Visit 6, Visit 7, Visit 8 and Visit 9 (end of treatment).

For each visit, ECG parameters are measured 3 times (5 minutes interval). For statistical analysis, the mean of those values will be taken into account.

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

A separate table will be produced, by treatment group and 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, flagged criteria and treatment group.

9.3.3 Physical Examination

A physical examination which includes body weight measurements will be performed at Visit 1 (screening), Visit 5, Visit 7 and Visit 9 (end of treatment). Height will be collected as part of the physical examination at Visit 1 (screening) and Visit 9 (end of treatment). If an accurate measurement of height is not possible an estimate will be provided.

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.

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 and treatment group.

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, by treatment group.

Change from 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 1 (screening) and Visit 9 (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 treatment group. 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 9 (end of treatment).

Menstruation details will be summarized as appropriate, including any changes in normal cycles at the end of treatment, by treatment group.

9.4 Other Measures

9.4.1 Concomitant Medication

Medications will be coded using the World Health Organization Drug Dictionary, Version B3 September 2018 or later version.

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. If a medication has a partial or missing start/stop date and it is unclear from the date whether the medication was taken after the first dose of IMP then it 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.

The start day and stop day will be included in the listing according to Section 6.1.5. If the date is partial and the exact day is unknown then the text ‘pre’ or ‘post’ will replace the start or stop day if it is clear from the partial date that the medication started or stopped prior to or after the first dose of IMP. If pre/post determination is not clear, Day will be left blank.

9.4.2 Procedures and Non-Drug Therapies

Procedures and non-drug therapies will be summarized by treatment group.

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 Plasma Concentrations of CBD and its Main Metabolites

The plasma concentrations of CBD and its main metabolites will be assessed at Visits 5, 6, 7 and 9 (if possible and as appropriate considering the patients weight, patient must weigh at least 12 kg). One trough PK blood sample will be taken at Visit 9 i.e., the sample must be collected before the patient’s next dose, 12 hours (−3/+6 hours) since last IMP dose. Visits where PK samples are taken will preferentially be scheduled for a time that allows collection of a trough PK sample,

whenever possible. The IMP can then be administered during the visit, once the PK sample has been collected.

Plasma concentrations of CBD and its main metabolites 7-hydroxy-CBD (7-OH-CBD) and 7-carboxy-CBD (7-COOH-CBD) will be summarized by visit and GWP42003-P arm, showing the number of non-missing values (n), arithmetic mean, standard deviation, coefficient of variation (%), median, minimum and maximum. Where samples are reported as BLQ, a value of zero will be used in the summaries. Summary statistics will be presented to 3 significant figures. Summaries by dosing method (oral/G tube) will be presented.

Plasma concentration summaries may exclude individual visits for patients deemed to meet certain criteria that could affect exposure. These criteria (identified in BDRM) include:

- Patients vomiting on or 1 day prior to the PK visit.
- Missed doses prior to the PK visit.
- IMP dose reduction.
- Cases of severe diarrhoea.
- Use of disallowed concomitant medication.

All data will be listed. Concentrations excluded from summaries will be flagged along with the reason for exclusion.

Patient meal times and details regarding the type of meal (snack, standard meal, high fat meal), and information on latest concomitant medicationwill be recorded at Visit 5, Visit 6, Visit 7 and Visit 9 (end of treatment). Meal and latest medication data will be listed only.**Blood Levels of Exploratory Biomarkers**

Biomarkers (e.g. IGF-1) and their change from baseline will be summarized by visit and treatment.

9.4.6 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 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 this study. As a result the analyses described in this SAP differ from those that were initially intended and presented in the protocol.

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.
6. Rubin, D. (1987). Multiple imputation for nonresponse in surveys. New York: John Wiley & Sons.

11. Tables, Listings, and Figures

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).

11.1 Planned Table Descriptions

Planned summary tables for protocol GWND18064 are fully detailed in the corresponding documentation on GWND 18064 TFL shells. Tables will be numbered according to the nomenclature used to support the clinical study report (CSR) according to sponsor's template.

11.2 Planned Listing Descriptions

Planned data and patient/patient data listings for protocol 4.1 are detailed in GW 18064 TFL shells document. Data listings will be numbered according to the nomenclature used to support the CSR when the Premier Research CSR template is used. However, the listing numbering structure can be modified per the sponsor's request to meet compatibility with the sponsor's CSR template.

In general, one listing will be produced per eCRF domain. All listings will be sorted by treatment, site, and patient number. All calculated variables will be included in the listings.

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.

11.3 Planned Figure

Planned figures for protocol 4.1 are detailed in GW 18064 TFL shells document.

12. Tables, Listings, and Listing Shells

12.1 Standard Layout for all Tables, Listings, and Figures

Table and listing 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.

No shells are provided for figures.

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
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
SAF	Safety population

Abbreviation	Definition
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/ECG

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 8.
For all classification below, a patient can fall in several categories for the same parameter (for example a patient with Diastolic BP > 100 will also be counted under Diastolic BP > 90).

Table 8 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 9.

Table 9 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

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 10.

Table 10 Defined Flagged Values for ECG Parameters

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

APPENDIX 4 Toxicity Criteria for Laboratory Parameters

The toxicity criteria that will be used to identify abnormal laboratory parameters are presented in Table 11 and Table 12.

Table 11 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 12 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 13 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 14 and Table 15.

Table 14 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 15 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 16.

Table 16 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 17 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 18.

Table 18 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 19.

Table 19 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 20.

Table 20 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 21.

Table 21 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 22.

Table 22 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 23.

Table 23 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.