Statistical Analysis Plan

A Randomized, Double-Blind, Placebo-Controlled Study of TEV-50717 (Deutetrabenazine) for the Treatment of Dyskinesia in Cerebral Palsy in Children and Adolescents

Study Number TV50717-CNS-30080

NCT03813238

SAP Approval Date: 04 May 2022

Statistical Analysis Plan Amendment 03

Study TV50717-CNS-30080

(RECLAIM-DCP)

A Randomized, Double-Blind, Placebo-Controlled Study of TEV-50717 (Deutetrabenazine) for the Treatment of Dyskinesia in Cerebral Palsy in Children and Adolescents

Short title: Reduction of Childhood and Adolescent Abnormal Involuntary Movements in Patients with Dyskinetic Cerebral Palsy in Children and Adolescents (RECLAIM-DCP)

Lay person title: A Study to Test if TEV-50717 is Effective in Relieving Abnormal Involuntary Movements in Cerebral Palsy

Phase 3

IND Number: 139700; NDA Number: NA; EudraCT Number: 2018-003742-17

Approval Date: 25 March 2019

Amendment 01 Approval Date: 23 March 2021

Amendment 02 Approval Date: 07 June 2021

Amendment 03 Approval Date: 04 May 2022

Sponsor

Teva Branded Pharmaceutical Products R&D, Inc.

145 Brandywine Parkway
West Chester, PA 19380
United States of America

Prepared by: [redacted]
Global Statistics, Teva

STATISTICAL ANALYSIS PLAN APPROVAL

Study No.: TV50717-CNS-30080

A Randomized, Double-Blind, Placebo-Controlled Study of TEV-50717 (Deutetrabenazine) for the Treatment of Dyskinesia in Cerebral Palsy in Children and Adolescents

Short title: Reduction of Childhood and Adolescent Abnormal Involuntary Movements in Patients with Dyskinetic Cerebral Palsy in Children and Adolescents (RECLAIM-DCP)

Statistical Analysis Plan for:

- ☒ **Interim Analysis**
- ☒ **Final Analysis**
- ☐ **Integrated Summary of Efficacy**
- ☐ **Integrated Summary of Safety**

Amendment: 03

Author: [redacted], **Teva Global Statistics**
[redacted], **Teva Global Statistics**

Executed signature pages are maintained separately within the Trial Master File

TABLE OF CONTENTS

List of Tables

LIST OF FIGURES

AMENDMENT HISTORY

The Statistical Analysis Plan for study TV50717-CNS-30080 (original study protocol dated 27 September 2018) has been amended and reissued as follows:

Amendment number	Date	Author(s)	Summary of changes	Reason for amendment
03	04 May 2022	[redacted], Teva Global Statistics [redacted], Teva Global Statistics	- Sample size reduction to approximately 65 patients due to slow enrollment - Remove the second interim analysis for futility and sample size reassessment - Add UHDRS-TMC and UHDRS-TMD (centrally read) for the scoring of chorea and dystonia, respectively, as key secondary endpoints - Add MCID and MCIC supportive primary analysis	Adjust the sample size and add new endpoints
02	06 June 2021	[redacted], Teva Global Statistics [redacted], Teva Global Statistics	The Wald statistic should be: $z_1 = \frac{-\hat{\delta}_1}{\mathrm{se}(\hat{\delta}_1)}$ Instead of what is currently written in the SAP: $z_1 = \frac{\lvert\hat{\delta}_1\rvert}{\mathrm{se}(\hat{\delta}_1)}$	Typo correction
01	23 March 2021	[redacted], Teva Global Statistics [redacted], Teva Global Statistics	- An additional interim analysis for futility when approximately 50 patients complete the study was added - A section for handling data from patients who terminate early from the study due to emergency circumstances (such as coronavirus disease 2019 [COVID-19] restrictions)	Update interim analyses based on protocol amendment 04

Amendment number	Date	Author(s)	Summary of changes	Reason for amendment
			and re-enter the study was added - Rules for missing items for MD-CRS I were clarified - Minor updates based on protocol amendment 04 - Editorial changes were made throughout the document for the purpose of clarification	
Original Statistical Analysis Plan	25 March 2019	[redacted], Teva Global Statistics [redacted], Teva Global Statistics	Not applicable	Not applicable

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Term
ADL	activities of daily living
BMI	body mass index
bpm	beats per minute
C-SSRS	Columbia-Suicide Severity Rating Scale
CaGI-I	Caregiver Global Impression of Improvement
CGI-I	Clinical Global Impression of Improvement
CGI-S	Clinical Global Impression of Severity
CI	confidence interval
CP	cerebral palsy
CRF	case report form
CSR	clinical study report
CTMS	Clinical Trial Management System
CYP2D6	cytochrome P450 2D6
DCP	dyskinesia in cerebral palsy
EAB	Enrollment Adjudication Board
ECG	Electrocardiogram/Electrocardiography
ET	early termination
FCS	fully conditional specification
HR	heart rate
iDMC	independent Data Monitoring Committee
IMP	investigational medicinal product
ITT	Intent-to-Treat
LS	least squares
MAR	missing at random
MCIC	minimal clinically important change
MCID	minimal clinically important difference
MD-CRS	Movement Disorder-Childhood Rating Scale
MD‑CRS I	Movement Disorder-Childhood Rating Scale Index I
MD‑CRS II	Movement Disorder-Childhood Rating Scale Index II

Abbreviation	Term
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified Intent-to-Treat
MMRM	mixed-effects model repeat measurement
MNAR	missing not at random
n	number
PEDI-CAT	Pediatric Evaluation Disability Inventory-Computer Adapted Test
PedsQL	Pediatric Quality of Life Inventory
PGI-I	Patient Global Impression of Improvement
PP	Per-Protocol
PR	Time from the beginning of the P wave to the beginning of the QRS complex in electrocardiogram
QRS	Graphical deflections corresponding to ventricular depolarization in a typical electrocardiogram
QTc	QT corrected for heart rate
QTcF	QT interval corrected for heart rate using Fridericia's formula
RR	Time between the start of one R wave and the start of the next R wave in the ECG
RTSM	Randomization and Trial Supply Management
SD	Standard Deviation
SDR	Statistical Data Review
SE	Standard Error
SOC	system organ class
SOP	standard operating procedure
UHDRS-TMC	Unified Huntington's Disease Rating Scale -Total Maximal Chorea
UHDRS-TMD	Unified Huntington's Disease Rating Scale -Total Maximal Dystonia
UHDRS-TMS	Unified Huntington's Disease Rating Scale - Total Motor Score
US	United States
WHO	World Health Organization
α-HTBZ	alpha-dihydrotetrabenazine
β-HTBZ	beta-dihydrotetrabenazine

INTRODUCTION

This Statistical Analysis Plan describes the planned analysis and reporting for Teva Branded Pharmaceutical Products R&D, Inc. Study TV50717-CNS-30080, A Randomized, Double-Blind, Placebo-Controlled Study of TEV-50717 (Deutetrabenazine) for the Treatment of Dyskinesia in Cerebral Palsy in Children and Adolescents, and was written in accordance with SOP GSD-SOP-702 (Statistical Analysis Plan).

The reader of this Statistical Analysis Plan is encouraged to read the study protocol for details on the conduct of this study, the operational aspects of clinical assessments, and the timing for completing the participation of a patient in this study.

The Statistical Analysis Plan is intended to be in agreement with the protocol, especially with regards to the primary and all secondary endpoints and their respective analyses. However, the Statistical Analysis Plan may contain more details regarding these particular points of interest, or other types of analyses (e.g. other endpoints).When differences exist in descriptions or explanations provided in the study protocol and this Statistical Analysis Plan, the Statistical Analysis Plan prevails; the differences will be explained in the Clinical Study Report (CSR).

1. STUDY OBJECTIVES

1.1. Primary Estimand

The primary estimand is the difference in means between TEV-50717 and placebo in the target patient population for the change from baseline to week 15 in centrally read Movement Disorder-Childhood Rating Scale (MD-CRS) part II total score, regardless of whether or not dose reduction, suspension, or discontinuation occurred, and regardless of treatment-related adverse events.

The primary estimand assesses the effectiveness in the reduction of choreiform movements in patients with dyskinesia in cerebral palsy (DCP) with predominant choreiform movement disorder, focusing on the causal effects attributable to the investigational medicinal product (IMP).

The target patient population for this study is patients with DCP with predominant choreiform movement disorder and severity of DCP represented by the following:

a. total score of ≥10 in centrally read MD-CRS part II items at baseline, and
b. Clinical Global Impression of Severity (CGI-S) ≥ 4 at baseline.

This population is expected to have the sensitivity to demonstrate clinically meaningful improvement following treatment with TEV-50717.

Due to practical reasons (periodical assessments), it is not possible to obtain a central reading of MD-CRS part II items at the time of the baseline visit prior to randomization; therefore, inclusion of the patient to the study is based on investigator scoring of MD-CRS part II items at baseline.

1.2. Secondary Estimands

The secondary estimands are the differences in means between TEV-50717 and placebo in the target patient population for:

(1) Change from baseline to week 15 in centrally read MD-CRS part I total score;

(2) Caregiver Global Impression of Improvement (CaGI-I) at week 15;

(3) Clinical Global Impression of Improvement (CGI-I) at week 15;

(4) Change from baseline to week 15 in centrally-read Unified Huntington's Disease Rating Scale –Total Maximal Chorea (UHDRS-TMC), and

(5) Change from baseline to week 15 in centrally-read Unified Huntington's Disease Rating Scale – Total Motor Dystonia (UHDRS-TMD)

Regardless of whether or not dose reduction, suspension, or discontinuation occurred, and regardless of treatment-related adverse events.

The secondary estimands assess the effectiveness on patients' ability to perform daily functions, the improvement in dyskinesia evaluated by the caregiver and the investigator, and the reduction

in chorea and dystonia evaluated by the central reviewers, all with a focus on the causal effects attributable to the IMP.

1.3. Primary and Secondary Study Objectives and Endpoints

The primary and secondary study objectives and endpoints are the following:

Objectives	Measures/Endpoints
The **primary objective** of the study is to evaluate the efficacy of TEV-50717 to reduce the severity of dyskinetic involuntary movements associated with CP.	The primary efficacy endpoint is the change from baseline to week 15 in the MD-CRS part II total score (movement disorder severity, centrally read) (TEV-50717 versus placebo).
A **secondary objective** is to evaluate the specific efficacy parameters of TEV-50717 beyond the measure of the primary objective.	The key secondary efficacy endpoints (TEV-50717 versus placebo) are the following: • MD-CRS part I total score (general assessment, centrally read) change from baseline to week 15 • CaGI-I (global, caregiver rated) at week 15 • CGI-I (global, physician rated) at week 15 • UHDRS-TMC (centrally read) at week 15 • UHDRS-TMD (centrally read) at week 15 Other efficacy measures and endpoints (TEV-50717 versus placebo) include the following: • MD-CRS part I total score (general assessment, physician rated) change from baseline to week 15 • MD-CRS part II total score (general assessment, physician rated) change from baseline to week 15 • MD-CRS Global Index (calculated from MD-CRS parts I and II total scores, centrally read) • MD-CRS Global Index (calculated from MD-CRS parts I and II total scores, physician rated) • UHDRS-TMS (physician rated) • UHDRS-TMC (physician rated) • UHDRS-TMD (physician rated) • PEDI-CAT (ADL, parent/caregiver completed, content-balanced version) • The CP module of the PedsQL (QoL,

Objectives	Measures/Endpoints
	patient/caregiver) • PGI-I (global, patient/caregiver) • CGI-S (global, physician rated) • CaGI-I response, defined as patients who are described by the caregiver as "Much Improved" or "Very Much Improved" in the CaGI-I score • CGI-I response, defined as patients who are described as "Much Improved" or "Very Much Improved" in the CGI-I score • CGI-S response, defined as patients who have a reduction of ≥1 point in the CGI-S score • PGI-I response, defined as patients who are described as "Much Improved" or "Somewhat Improved" in the PGI-I score
A **secondary objective** of the study is to evaluate the safety and tolerability of TEV-50717.	The safety variables include adverse events (and the number of patients who withdraw from the study due to adverse events), vital signs, laboratory tests (hematology, chemistry, and urinalysis), ECG measurements, CBCL, ESRS, ESS, and the children's C-SSRS.

ADL=activities of daily living; CaGI-I=Caregiver Global Impression of Improvement; CBCL=Child Behavior Checklist (for ages 6-18); CGI-I=Clinical Global Impression of Improvement; CGI-S=Clinical Global Impression of Severity; CP=cerebral palsy; C-SSRS=Columbia-Suicide Severity Rating Scale; ECG=electrocardiogram; ESRS=Extrapyramidal Symptom Rating Scale (subscales I and II); ESS=Epworth Sleepiness Scale (for children and adolescents); MD-CRS=Movement Disorder-Childhood Rating Scale; PEDI-CAT=Pediatric Evaluation Disability Inventory-Computer Adapted Test; PedsQL=Pediatric Quality of Life Inventory; PGI-I=Patient Global Impression of Improvement; QoL=quality of life; UHDRS-TMC=Unified Huntington's Disease Rating Scale-Total Maximal Chorea; UHDRS-TMD=Unified Huntington's Disease Rating Scale-Total Maximal Dystonia; UHDRS-TMS=Unified Huntington's Disease Rating Scale-Total Motor Score.

1.4. Exploratory Objectives

[redacted]

2. STUDY DESIGN

2.1. General Design

This is a Phase 3, 21-week, multicenter, randomized, double-blind, placebo-controlled, parallel group study to evaluate the efficacy and safety of TEV-50717 administered as oral tablets at a starting dose of 6 mg once daily in patients (age 6 through 18 years, inclusive) with DCP with predominant choreiform movement disorder, who have had cerebral palsy (CP) symptoms, of a non-progressive nature, since infancy (≤2 years of age).

The study will consist of a screening period (up to 31 days) and a double-blind treatment period including a titration period (7 weeks) and a maintenance period (8 weeks), followed by a washout period of 1 week, and a follow-up telephone contact 1 week after the washout period.

At the baseline visit (day 1), patients will be randomly assigned to 1 of 2 treatment groups with TEV-50717 IMP or placebo IMP in a 2:1 ratio stratified by age at baseline (6 to <12 years; 12 through 18 years, inclusive) and region (US; non-US).

Titration period (7 weeks): The titration scheme and maximum dose will be determined by body weight and cytochrome P450 2D6 (CYP2D6) impairment status at baseline[1]; patients will be classified as CYP2D6 impaired if they are receiving a strong CYP2D6 inhibitor or are a CYP2D6 poor metabolizer. IMP will be titrated during the double-blind treatment period starting with 6 mg of TEV-50717 or matching placebo IMP. During the titration period, patients and their caregiver/adult will interact weekly with the investigator/staff from week 1 through week 7: in-person (in-clinic) study visits will be scheduled at weeks 3 and 7, and telephone contacts (with non-recording live video) will be scheduled for weeks 1, 2, 4, 5, and 6 in order to assess dyskinesia and adverse events. The telephone contacts to the patient will be supported by live video stream, without recording, to provide visual confirmation to the investigator of the verbal information provided by the patient or caregiver. The dose of IMP will be adjusted according to the titration scheme to identify a dose level that optimally reduces dyskinesia (as determined by the investigator) and is well tolerated. If a patient experiences a "clinically significant" adverse event attributable to the IMP, the investigator will first determine if a dose reduction (to the previous dose level) or suspension is necessary and possible.

Maintenance period (8 weeks): Patients will continue to receive their maintenance dose over the next 8 weeks, although a 1-time dose reduction (to the previous dose level for the remainder of the study) for adverse events is allowed. Patients will return to the clinic at weeks 9, 12, and 15 for assessments of efficacy and safety.

Washout period and follow-up: All patients will discontinue IMP at the week 15 visit and will return 1 week later for the end-of-study (EOS) visit. Patients who complete the study may be eligible to begin participation in the open label safety extension Study TV50717-CNS-30081 at that time. At the week 16 visit, patients may choose to enter Study TV50717-CNS-30081 (on

[1] The titration scheme is presented in Tables 4 and 5 in the study protocol.

that day) or up to 1 week later (2 weeks from the week 15 visit). Patients who complete treatment but choose not to participate in the open-label safety extension Study TV50717-CNS-30081 before week 17 of Study TV50717-CNS-30080 will be contacted for a safety evaluation call occurring 2 weeks after their last dose of IMP.

The study scheme is presented in Figure 1.

At the time of informed consent, the parent/legal guardian will be counseled that, once randomized to treatment, patients are to remain in the study and complete all study procedures unless the choice is made to withdraw consent. Patients who withdraw from the study before completing the 15 week treatment period should have an early termination (ET) visit as soon as possible after the last dose of IMP. All patients who discontinue early will have a follow-up telephone contact for safety evaluation 2 weeks after ET; evaluations will be as described for week 17.

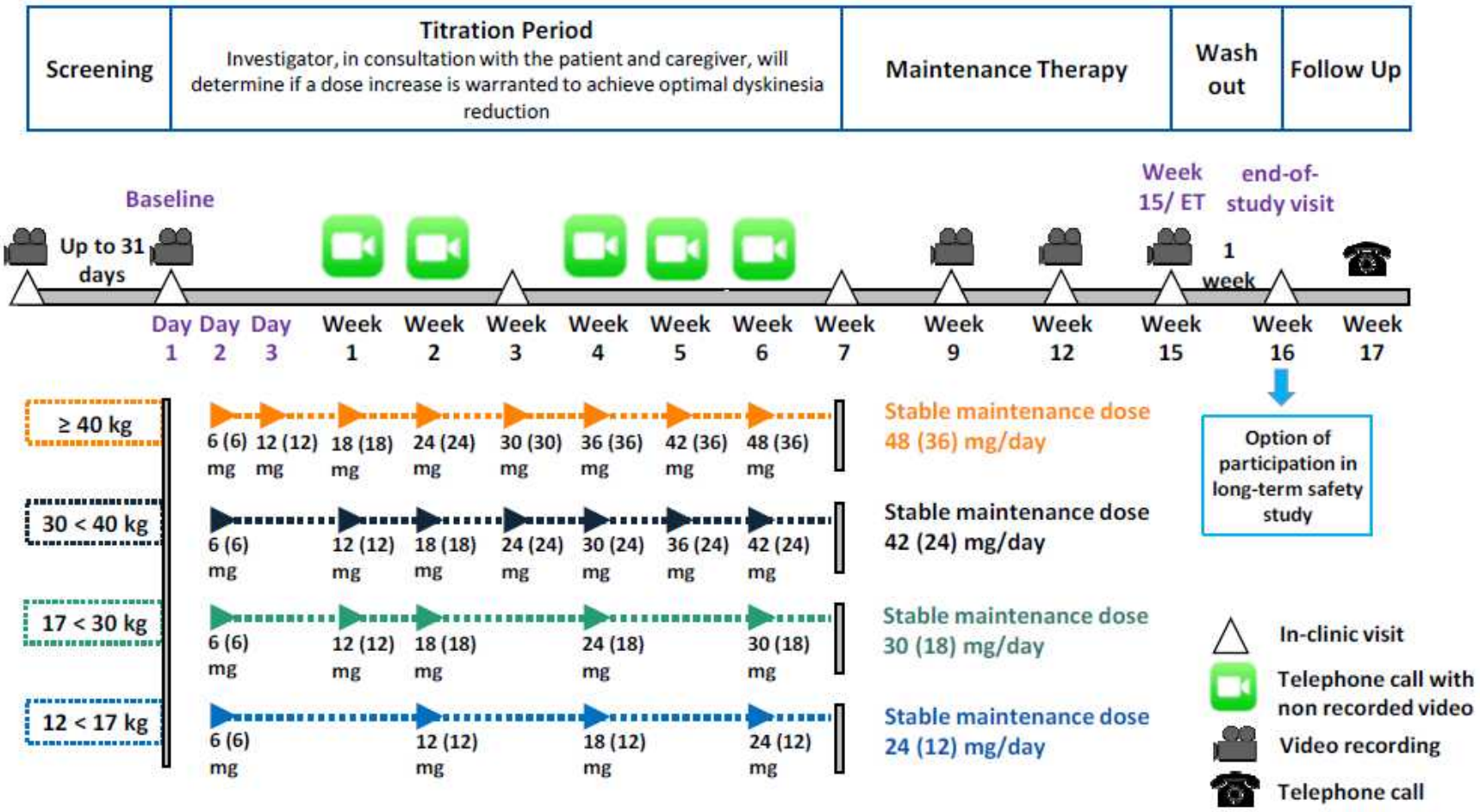
Patients who screen fail or terminate early from the study due to emergency circumstances (such as coronavirus disease 2019 [COVID-19]) may be allowed to re-screen/re-enter (respectively) the study at a later date, depending on the circumstances.

Study procedures and assessments with their timing are summarized in Table 3 of the study protocol.

Interim Analysis: When approximately 50 patients have completed the study (including follow-up), an independent Data Monitoring Committee (iDMC) will perform an unblinded non-binding IA for futility based on centrally read MD-CRS part II total score.

An iDMC charter will be developed for the IA, and the procedures to ensure the integrity of the study will be provided in the charter, see details in Section 10.

Figure 1: Overall Study Schematic Diagram



AE=adverse event; CGI=Clinical Global Impression; CYP2D6=cytochrome P450 2D6; ET=early termination; max=maximum.
Note: If a patient is CYP2D6-impaired, the dose administered is indicated in parentheses. The dose of the IMP should be increased on a weekly basis until any of the following events occur: clinically meaningful reduction in dyskinesia (ie, CGI-I) as determined by the investigator, clinically significant AE, or the maximum allowable dose is reached.

2.2. Randomization and Blinding

This is a double-blind study. Patients will be randomized to either TEV 50717 or matching placebo in a 2:1 ratio using Randomization and Trial Supply Management (RTSM). Patients will be stratified by age at baseline (6 to <12 years; 12 through 18 years, inclusive) and region (US; non-US). Patients and investigators will remain blinded to treatment assignment during the study. In addition, the sponsor's clinical personnel and all vendors (with the exception of the RTSM vendor) will be blinded to the IMP identity until the database has been locked for analysis and the treatment assignments are revealed.

Patients re-entering the study will be manually assigned to their initial treatment group by the RTSM unblinded vendor. Blinding will be maintained for the patient, investigator, and sponsor as described above; the only information that will be available is that the patient was assigned to the same treatment group as previously.

The iDMC will perform an unblinded non-binding IA for futility and will monitor safety throughout the study based on unblinded data. An iDMC charter will be developed for the IA, and the procedures to ensure the integrity of the study will be provided in the charter.

In the event of an emergency, it will be possible to determine to which treatment group and dose the patient has been allocated by accessing the RTSM system. All investigational centers will be provided with details of how to access the system for code breaking at the start of the study. The medical monitor or equivalent should be notified following unblinding. Any unblinding of the IMP performed by the investigator must be recorded in the source documents.

The randomization list will be assigned to the relevant treatment groups through a qualified service provider (eg, via an RTSM system). The generation of the randomization list and management of the RTSM system will be done by a qualified service provider under the oversight of the responsible function at Teva.

Patient randomization codes will be maintained in a secure location at the service provider contracted to generate the codes.

At the time of IA analysis, after receiving unblinding request from a Teva statistician, the service provider will provide the unblinded treatment assignment to an independent statistician according to the processes defined in the relevant standard operating procedure. The independent statistician will perform the IA and prepare unblinded outputs for the iDMC safety monitoring sessions. The treatment codes should remain undisclosed to the sponsor in order to ensure the integrity of the ongoing study conduct.

At the time of final analysis, after receiving unblinding request from a Teva statistician, the service provider will provide the Teva statistician with the unblinded IMP assignment according to the processes defined in the relevant standard operating procedure.

2.3. Data Monitoring Committee

There will be an iDMC in this study. When approximately 50 patients have completed the study (including follow-up), the iDMC will perform an unblinded non-binding IA for futility based on centrally read MD-CRS part II total score (see Section 10).

The iDMC will also perform regular safety monitoring throughout the study.

Details on how the iDMC chairperson will communicate with the sponsor in regard to issues resulting from the conduct and clinical aspects of the study are described in the iDMC charter. The iDMC charter will be developed for the IA, and the procedures to ensure the integrity of the study will be provided in the charter. The sponsor will work closely with the iDMC to provide the necessary safety and efficacy data for review. In addition, results of the IA will be communicated to the sponsor as outlined in the iDMC charter.

2.4. Sample Size and Power Considerations

The initial sample size estimation of approximately 185 patients randomized to TEV-50717 versus placebo in a 2:1 ratio (approximately 124 in the TEV-50717 group; approximately 61 in the placebo group) is the sample size required to obtain a statistically significant result in the primary analysis based on a 1-sided test for mean difference at a significance level of α=0.025 with a power of approximately 90%, assuming a mean difference of 1.8 and a standard deviation (SD) of 3.5 in each treatment group.

Based on Protocol Amendment 05, the initial planned sample size has been reduced to approximately 65 patients randomized to TEV-50717 versus placebo in a 2:1 ratio (approximately 43 in the TEV-50717 group; approximately 22 in the placebo group).

The initial assumptions regarding the mean difference and SD are based on the change from baseline to month 6 in patients treated with trihexyphenidyl reported by Battini et al (2014), assuming an effect size of 55% for TEV-50717 versus placebo. The effect size of 55% was assumed based on the observed effect sizes of TEV-50717 in the pivotal studies in Huntington's disease and tardive dyskinesia. In the current study, these assumptions correspond to a mean difference of 2.0 in MD-CRS part II total score and an SD of 3.5 in each treatment group.

The rate of missing not at random (MNAR) data for the primary analysis is expected to be 10%. MNAR data will be imputed in a conservative manner as if they were obtained from patients in the placebo treatment group, regardless of the actual treatment group. As a result, the mean difference is expected to be reduced from 2.0 to 1.8 in the primary analysis. The impact of the missing data imputation on the SD is negligible.

2.5. Sequence of Planned Analyses

2.5.1. Planned Interim Analyses

This is the first study to evaluate the efficacy of TEV-50717 to reduce the severity of dyskinetic involuntary movements associated with CP in pediatric patients with DCP. Moreover, limited data are available on the effect of active treatment on MD-CRS part II total score in DCP that can support sample size calculations. Therefore, an IA for futility is planned.

When approximately 50 patients have completed the study (including follow-up), an iDMC will perform an unblinded non-binding IA for futility based on centrally read MD-CRS part II total score.

The rules for futility will be included in the iDMC charter. The iDMC charter will also provide the wording that will be used to communicate the iDMC recommendation.

The Statistical Analysis Plan will be approved before the IA will be performed, in accordance to SOP GBP_RD_702 (Statistical Analysis Plan). Any changes in the main analysis of the primary and key secondary endpoints (eg, model specifications) will not be allowed after the IA.

2.5.2. Final Analyses and Reporting

All analyses identified in this Statistical Analysis Plan will be performed after the end of study as defined in the study protocol.

This final version of the Statistical Analysis Plan and any corresponding amendments will be approved before database lock, in accordance to SOP GBP_RD_702 (Statistical Analysis Plan).

The randomization codes will not be unblinded until this Statistical Analysis Plan has been approved and issued.

3. ANALYSIS SETS

3.1. Intent-to-Treat Analysis Set

The intent to treat (ITT) analysis set will include all randomized patients.

In the ITT analysis set, treatment will be assigned based on the treatment to which patients were randomized, regardless of which treatment they actually received.

For patients who discontinue the study and re-enter it (repeating the study from baseline; see Section 4.5), only the data from the repeated participation will be included in the ITT analysis set.

3.2. Modified Intent-to-Treat Analysis Set

The modified intent-to-treat (mITT) analysis set is a subset of the ITT analysis set including all randomized patients with at least 1 post-baseline centrally read MD-CRS part II assessment.

In the mITT analysis set, treatment will be assigned based on the treatment to which patients were randomized, regardless of which treatment they actually received.

3.3. Safety Analysis Set

The safety analysis set will include all randomized patients who receive at least 1 dose of IMP.

In the safety analysis set, treatment will be assigned based on the treatment patients actually received, regardless of the treatment to which they were randomized, unless otherwise specified.

Rules for assignment of treatment in case of mixed actual treatments:

- If a patient received placebo throughout the entire treatment period, then actual treatment is placebo.
- If a patient was randomized to placebo and received TEV-50717 for up to 7 days, then actual treatment is placebo
- Otherwise, actual treatment is TEV-50717.

For patients who discontinue the study and re-enter it (repeating the study from baseline; see Section 4.5) only the data from the repeated participation will be included in the safety analysis set.

3.4. Per-Protocol Analysis Set

The per-protocol (PP) analysis set is a subset of the mITT analysis set including only patients who meet the following criteria:

- compliant with study medication (80% to 105%)
- complete the study without any major protocol deviations that may impact efficacy
- centrally read baseline MD-CRS part II total score ≥10

Examples of major protocol deviations that may result in patient exclusion from the PP analysis set are given in Table 1.

Table 1: Possible Types of Major Protocol Deviations that May Lead to Exclusion from the Per-Protocol Analysis Set

Type of deviation	Description of Deviation
Informed consent	Missing informed consent
Inclusion/Exclusion	Patient did not meet inclusion/exclusion criteria
Compliance	Overall compliance during the treatment period <80% or >105%
	Overall compliance during the maintenance period <80% or >105%
Study drug	Patient took incorrect dose for more than 7 days
Dosing error	Patient received the incorrect IMP, contrary to the randomization schedule (eg, patient randomized to placebo group and received a supply of TEV-50717).
Restricted/disallowed Concomitant Medication	Patient took a medication that was not permitted during participation in this study
MD-CRS part II total score	Missing centrally-read MD-CRS part II assessment at baseline
	Patient does not have at least 1 post-baseline centrally read MD-CRS part II assessment

Evaluation of exclusion from the PP analysis set will be discussed on a case-by-case basis and documented prior to database lock and unblinding of the study for final analysis.

3.5. Modified Intent-to-Treat Analysis A Set

The modified intent-to-treat A (mITTA) analysis set is a subset of the mITT analysis set including all patients that received at least 1 dose of IMP.

In the mITTA analysis set, treatment will be assigned based on the treatment to which patients were randomized, regardless of which treatment they actually received.

3.6. Modified Intent-to-Treat Analysis B Set

The modified intent-to-treat B (mITTB) analysis set is a subset of the mITT analysis set including all patients that received at least 1 dose of IMP and have centrally-read MD-CRS part II total score ≥10 at baseline.

In the mITTB analysis set, treatment will be assigned based on the treatment to which patients were randomized, regardless of which treatment they actually received.

4. GENERAL ISSUES FOR DATA ANALYSIS

4.1. General

Descriptive statistics for continuous variables include number (n), mean, standard deviation (SD), standard error (SE), median, minimum, and maximum. Mean, SD and SE will not be presented if the number of patients in the analysis is less than 5 ($n < 5$).

Descriptive statistics for categorical variables include patient counts and percentages. Missing category will be displayed as appropriate.

The following variables are used as stratification factors for randomization or determination of titration scheme and maximal dose: CYP2D6 impairment status; weight category; age group; region. For the purpose of analysis, the actual baseline values as derived from the Case Report Form (CRF), or from the CRF and the laboratory database in case of CYP2D6 impairment status, will be used in the case of discrepancy between Randomization and Trial Supply Management (RTSM) and CRF data.

4.2. Specification of Baseline Values

Generally, the baseline value is the last observed data before the first dose of IMP, unless otherwise noted. If the baseline assessments occur on the same day as the first dose of IMP, it is assumed that the patients follow study procedure and those baseline assessments will be used. For patients in the ITT population who took no IMP, the baseline value is the last observed data on or prior to the randomization date. Electrocardiogram/Electrocardiography (ECG) data will follow this rule, unless specified otherwise.

MD-CRS part I and II central reading: for specification of baseline values in case of missing centrally-read MD-CRS part I or II at baseline, see Sections 4.3.1 and 4.3.2.

Baseline values for patients withdrawals and re-enter the study will be defined in Section 4.5.

4.3. Handling Withdrawals and Missing Data

For this study design and population, it is unlikely that patients who withdraw from IMP will be willing to attend the remaining scheduled study visits and, in particular, the week 15 visit. Therefore, patients will not be required to return to the clinic at week 15 for efficacy evaluations, and the primary analysis of the primary and key secondary efficacy endpoints will be based on a multiple imputation method for missing data.

For the primary efficacy and key secondary analyses, missing data will be classified as missing at random (MAR) and MNAR. The following cases will be considered as MNAR:

- early terminations related to tolerability
- early terminations related to treatment-related adverse events
- early terminations related to lack of efficacy

The following cases will be considered as MAR:

- intermittent missing data

- early termination for patients who are lost to follow-up
- early termination due to COVID-19

For all other cases, and for cases classified as above, classification as MAR/MNAR will be done/revisited in a blinded manner on a case-by-case basis prior to the IA for futility and prior to database lock for final analysis and will be documented in a dedicated document.

Multiple imputation methods will be used in the primary and secondary efficacy analyses. See Sections 6.2 - 6.3 for details.

Note: imputation of missing baseline values will be performed prior to implementation of multiple imputations in the analysis model.

4.3.1. Specification of Baseline Value for Centrally-Read MD-CRS II

- General rule: baseline assessment (central reading)
- If above not available, then Enrollment Adjudication Board (EAB) screening assessment;
- If above not available, then impute from the local reading of baseline using Regression imputation (see details below);
- If above not available, then the average of the centrally-read baseline MD-CRS part II total score across all patients from the ITT analysis set (Section 3.1) with baseline value.

Regression imputation is a single imputation method. Let Y_1 be the local reading of baseline MD-CRS II total score; Y_2 be the centrally-read MD-CRS II. On the observed pairs (Y_1, Y_2), fit Y_2 based on Y_1 as $\widehat{Y_2} = Y_1\beta + \alpha$ using linear regression. Impute the missing Y_2 values using the predicted value for a given Y_1.

4.3.2. Specification of Baseline Value for Centrally-Read MD-CRS I

- General rule: baseline assessment (central reading);
- If above not available, then impute from the local reading of baseline using Regression Imputation (see Section 4.3.1 for details);
- If above not available, then the average of the centrally-read baseline MD-CRS part I total score across all patients from the ITT analysis set (Section 3.1) with baseline value.

4.3.3. Specification of Baseline Value For Clinical Global Impression of Severity

If the baseline assessment of the Clinical Global Impression of Severity (CGI-S) is not available, then the average of the baseline CGI-S assessments across all patients from the ITT analysis set (Section 3.1) with baseline value will be used instead. Note that this imputation rule will only be applied for the analysis of the Clinical Global Impression of Improvement (CGI-I).

4.3.4. Specification of Baseline Value for Centrally-Read UHDRS-TMC and TMD

- General rule: baseline assessment (central reading);

- If above not available, then impute from the local reading of baseline using Regression Imputation (see Section 4.3.1 for details);
- If above not available, then the average of the centrally-read baseline value across all patients from the ITT analysis set (Section 3.1) with baseline value.

4.3.5. Missing Items in Rating Scales

4.3.5.1. Missing Items in MD-CRS

The MD-CRS has 2 parts: part I for general assessment of the patient's functioning and impact of cerebral palsy on the activities of the patient, and part II for a specific motor assessment of the severity of the movement disorder.

There are 15 items in MD-CRS Part I. Items A1-A7 (Motor function), C1 (Self-dressing) and D1 (Attention/alertness during the observation) are scored by the rater in the clinic and these are video recorded for central reading, all other Part I items are scored by the rater based on the parent/caregiver interview. The centrally-read MD-CRS Part I total score is obtained by summing the centrally-read items and the items that are scored only by the rater in the clinic. Each item is on a 0 (present) to 4 (absent) scale. The minimum score is 0 and the maximum score is 60.

There are 7 items in MD-CRS Part II. All items are scored by the rater in the clinic and are centrally read based on video recording. The minimum score is 0 (absent) and the maximum score is 28 (marked/prolonged).

The MD-CRS part I total score: in case of any missing items, the total score will be derived by calculating the MD-CRS I index (see Section 6.5.1.1), and multiplying by the maximum score of 60.

The MD-CRS part II total score: in case of any missing items, the total score will be derived by calculating the MD-CRS II index (see Section 6.5.1.1), and multiplying by the maximum score of 28.

Missing items will be discussed in the blinded statistical data review (SDR) meeting prior to database lock.

4.3.5.2. Other Missing Items in Rating Scales

Rules for handling missing items in efficacy and safety rating scales are presented in Section 6.4 (Unified Huntington's Disease Rating Scale-Total Motor Score (UHDRS-TMS), Unified Huntington's Disease Rating Scale-Total Maximal Chorea (UHDRS-TMC), Unified Huntington's Disease Rating Scale-Total Maximal Dystonia (UHDRS-TMD), Pediatric Evaluation Disability Inventory-Computer Adapted Test (PEDI-CAT), Pediatric Quality of Life Inventory (PedsQL) and Patient Global Impression of Improvement (PGI-I)) and Section 8 (Columbia-Suicide Severity Rating Scale [C-SSRS], Child Behavior Checklist [CBCL], Extrapyramidal Symptom Rating Scale [ESRS] and Epworth Sleepiness Scale [ESS]).

Unless otherwise specified, only those observed data from the patients will be used in the statistical analyses, ie, no estimation of missing data.

4.4. Study Days and Visits

Study days will be numbered relative to the baseline visit date. Day 1 is defined as the date of the baseline visit, and the first dose of the IMP will be taken the next morning (day 2). Days will be numbered relative to baseline (ie, ..., –2, –1, 1, 2, ...; day –1 being the day before the baseline visit).

For efficacy and safety by-visit analyses, data collected at post-baseline scheduled visits will be included using their scheduled visit.

Data collected at early termination and unscheduled visits will be included and assigned to a visit window as described below.

Table 2: Visit Windows for Early Termination Visit for Efficacy Assessments

Assessment[a]	Study Day Window	Scheduled Day	Scheduled Visit/Week
MD-CRS	Day 2 - 73	Day 63	Week 9
	Day 74 - 94	Day 84	Week 12
	Day 95 - EOT	Day 105	Week 15
	≥Day after EOT	Day 112	Week 16
CaGI-I, CGI-S, PGI-I	Day 2 - 35	Day 21	Week 3
	Day 36 - 53	Day 49	Week 7
	Day 54 - 73	Day 63	Week 9
	Day 74 - 94	Day 84	Week 12
	Day 95 - EOT	Day 105	Week 15
	≥Day after EOT	Day 112	Week 16
CGI-I	Day 2 - 10	Day 7	Week 1
	Day 11 - 17	Day 14	Week 2
	Day 18 – 24	Day 21	Week 3
	Day 25 - 31	Day 28	Week 4
	Day 32 - 38	Day 35	Week 5
	Day 39 - 45	Day 42	Week 6
	Day 46 - 53	Day 49	Week 7
	Day 54 - 73	Day 63	Week 9
	Day 74 - 94	Day 84	Week 12
	Day 95 - EOT	Day 105	Week 15
	≥Day after EOT	Day 112	Week 16
PEDI-CAT, PedsQL	Day 2 – EOT	Day 105	Week 15

Assessment[a]	Study Day Window	Scheduled Day	Scheduled Visit/Week
UHDRS-TMS, UHDRS-TMC, UHDRS-TMD	Day 2 - 73	Day 63	Week 9
	Day 74 - 94	Day 84	Week 12
	≥Day 95 - EOT	Day 105	Week 15

[a] CaGI-I=Caregiver Global Impression of Improvement; CGI-I=Clinical Global Impression of Improvement; CGI-S=Clinical Global Impression of Severity; PGI-I=Patient Global Impression of Improvement; UHDRS-TMS=Unified Huntington's Disease Rating Scale-Total Motor Score; UHDRS-TMC=Unified Huntington's Disease Rating Scale-Total Maximal Chorea; UHDRS-TMD=Unified Huntington's Disease Rating Scale-Total Maximal Dystonia; EOT=end of treatment

Table 3: Visit Windows for Early Termination Visit for Safety Assessments

Assessment[a]	Study Day Window	Scheduled day	Scheduled Visit/Week
Vital signs, Weight, C-SSRS, ESRS, CBCL- Syndrome Scale, ESS	Day 2 - 35	Day 21	Week 3
	Day 36 - 53	Day 49	Week 7
	Day 54 - 74	Day 63	Week 9
	Day 75 - 94	Day 84	Week 12
	Day 95 – EOT	Day 105	Week 15
	≥Day after EOT	Day 112	Week 16
ECG	Day 2 - 35	Day 21	Week 3
	Day 36 - 67	Day 49	Week 7
	Day 68 – EOT	Day 105	Week 15
	≥Day after EOT	Day 112	Week 16
Laboratory tests other than pregnancy and drug screen	Day 2 – EOT	Day 105	Week 15
	≥Day after EOT	Day 112	Week 16
Orthostatic blood pressure, Physical exam, Neurological exam, height, CBCL-full assessment	Day 2 – EOT	Day 105	Week 15

[a] CBCL=Child Behavior Checklist (for ages 6-18); C-SSRS=Columbia-Suicide Severity Rating Scale; ECG=electrocardiogram; ESRS=Extrapyramidal Symptom Rating Scale (subscales I and II); ESS =Epworth Sleepiness Scale (for children and adolescents); EOT=end of treatment

After mapping the data to the analysis visits of unscheduled and early termination visits, the following rules will apply unless other handling is specified for a particular analysis:

- If there is a scheduled visit in the analysis window, the last non- missing scheduled visit will be selected.
- If there is no scheduled visit, but early termination visit, early termination visit will be selected.

- If there are no scheduled or early termination visits, the record closest to the planned assessment day will be selected for analysis.
- If there are no scheduled or early termination visits, and 2 records are equidistant from the scheduled day, then the later record will be selected.
- If a patient has no scheduled record or early termination/unscheduled visits in an analysis window, the patient will be considered missing at that visit.

For these analyses performed by time period, eg, adverse events, the following definitions will be used:

- Titration Period - from the first dose of treatment until the day of the week 7 visit. For patients discontinued before the week 7, the Titration Period is until the later of the last dose of treatment and the early termination visit.
- Maintenance Period - from the day after the week 7 visit until the day of the week 15 visit. If a patient enters the Maintenance Period but discontinues treatment early, the Maintenance Period is until the later of the last dose of treatment and the early termination visit.
- Overall Treatment Period - includes the Titration and the Maintenance Periods, if applicable.
- Follow-up Period - from the end of the Overall Treatment Period through the end of the study.

For last assessment, the last available non-missing value will be used. Unscheduled visits that could not be mapped will not be displayed in the safety by-visit summaries, but they will be considered for the endpoint/last assessment visit.

4.5. Handling Data From Patients Withdraw and Re-Enter the Study

Patients who terminate early from the study due to emergency circumstances (such as coronavirus disease 2019 [COVID-19]) may be allowed to re-enter the study at a later date. Patients re-entering the study will be manually assigned to their initial treatment group and will redo the entire study from the start.

4.5.1. Baseline Values

In general, baseline values from the re-entered participation will be used. Missing baseline values will be handled in the same method as other non-withdrawal patients.

The blinded CYP2D6 genotyping is the only measurement that is not required for re-entering patients. Results from the original participation of the CYP2D6 impairment status at baseline will be carried over to the repeated participation (re-entered) portion.

4.5.2. Efficacy Data

For each patient that re-entered the study, only data from the repeated participation will be included in the ITT, mITT, PP, mITTA, and mITTB analysis sets. The original participation data will be listed separately and will not be included in any summary tables or statistical analysis.

Missing postbaseline efficacy values will be handled in the same method as other non-withdrawal patients.

4.5.3. Safety Data

For each patient that re-entered the study, only data from the repeated participation will be included in the ITT and safety analysis sets. The original participation data will be listed separately and will not be included in any summary tables. In case of a SAE that occurred during the original participation and resolved prior to re-entering, the SAE will be documented in medical history database. Missing postbaseline safety values will be handled in the same method as other non-withdrawal patients.

5. STUDY POPULATION

5.1. General

The ITT analysis set (Section 3.1) will be used for all study population summaries, unless otherwise specified. The summaries might be presented using the mITT analysis set (Section 3.2), in case >10% of ITT patients were excluded from the mITT analysis set. Summaries will be presented by treatment group and for all patients.

5.2. Patient Disposition

Data from patients screened; patients screened but not randomized (and reason not randomized); patients who are randomized; patients randomized but not treated; patients re-randomized (if applicable); patients in the ITT, mITT, safety, PP, mITTA and mITTB analysis sets; patients who complete the treatment period; patients who withdraw from the study during the treatment period; patients who complete the EOS visit; and patients who withdraw from the study will be summarized using descriptive statistics.

Data from patients who withdraw from the study during the treatment period, and from the study overall will be summarized by reason for withdrawal.

Data from patients who withdraw during the treatment period will be summarized by classification to MAR/MNAR overall, for the titration and maintenance periods, and by visit using descriptive statistics.

5.3. Demographics and Baseline Characteristics

Patient demographic and baseline characteristics will be summarized using descriptive statistics.

The continuous variables of patient age, weight, height, and body mass index (BMI) will be summarized using descriptive statistics.

BMI at baseline will be computed as:

BMI (kg/m^2) = Weight at baseline (kg)/[Height at screening $(m)^2$].

Height, weight, and BMI at baseline will also be presented by age group and sex using descriptive statistics. In addition, normal age and sex-based z-scores and percentiles for BMI will be determined using the World Health Organization (WHO) growth charts, and summarized descriptively.

The categorical variables of patient weight group, sex, race, ethnicity, and region (United States (US) and non- US) and country for non-US will be summarized using descriptive statistics for each category. The categorical variables of age group at baseline (6 to <12 years, 12 through 18 years, inclusive), use of a strong CYP2D6 inhibitor, as appears in Table 9 of the study protocol (Yes/No), CYP2D6 genotype (Poor CYP2D6 metabolizer/ Non-poor CYP2D6 metabolizer), and CYP2D6 impairment status (Impaired/Not impaired) will be summarized using descriptive statistics for each category. In addition, age and sex-based BMI categories include: Underweight (< 5 percentile), Normal (≥ 5 - < 85 percentile), Overweight (≥ 85 - < 95 percentile), Obese (≥ 95 percentile), will be tabulated for each category.

5.4. Disease Characteristics

Post-hoc analysis will be provided for disease characteristics.

5.5. Medical History

All medical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The incidence of medical history abnormalities will be summarized using descriptive statistics by system organ class (SOC) and preferred term. Patients are counted only once in each preferred term and SOC category. Summaries will be presented by treatment group and for all patients.

5.6. Prior Therapy and Medication

Any prior or concomitant medication a patient has had within 3 months prior to baseline and up to the end of study, including follow up, will be recorded on the CRF. Generic or trade name, indication, dose, and start and end dates of the administered medication will be recorded. The sponsor will encode all medication according to the WHO drug dictionary.

The incidence of prior therapies and medications will be summarized using descriptive statistics by therapeutic class, preferred base name and preferred term. Patients are counted only once in each therapeutic class category, and only once in each preferred term category. Prior therapies and medications will include all medications taken and therapies administered before the first day of IMP administration.

5.7. Study Drug Titration

The number and percent of patients at each dose level by weight group and CYP2D6 impairment status during the titration period will be summarized by week. The latest administered dose in each study week will be used.

Reasons for the deviation from the IMP dosing scheme (see Protocol Sections 5.1.1.1 and 5.1.1.2) will be summarized overall and by period (titration and maintenance). Patient data will be reviewed in the blinded SDR meeting prior to database lock to determine the reason for deviation in each period and overall.

Descriptive statistics of the maintenance dose (last dose taken in the maintenance period); nominal and as percent of the maximal planned dose for the patient) will be presented.

The safety analysis set will be used for IMP titration summaries.

5.8. Study Drug Compliance

Study drug compliance during the treatment period will be assessed as described in the protocol Section 5.7.

Study drug compliance will be determined by presenting the number of tablets taken as a percentage of the expected number of tablets based on the dose(s) of the patient at the relevant time period (see Table 4 and Table 5; study protocol Section 5.1).

Table 4: Drug Kits based on Titration Scheme – Non-Impaired CYP2D6 Metabolizers

Weight Group	12kg to <17kg		17kg to <30kg		30kg to <40kg		≥40kg	
Visit	Dose (mg/day)	Tablets (mg)	Dose (mg/day)	Tablets (mg)	Dose (mg/day)	Tablets (mg)	Dose (mg/day)	Tablets (mg)
Day 2	6	6	6	6	6	6	6	6
Week 1	6	6	6	6	6	6	12	6,6
Week 2	6	6	12	6,6	12	6,6	18	9,9
Week 3	12	6,6	18	9,9	18	9,9	24	12,12
Week 4	12	6,6	18	9,9	24	12,12	30	15,15
Week 5	18	9,9	24	12,12	30	15,15	36	18,18
Week 6	18	9,9	24	12,12	36	18,18	42	15,6 AM 15,6 PM
Week 7	24	12,12	30	15,15	42	15,6 AM 15,6 PM	48	12,12 AM 12,12 PM

Dose and tablets refer to IMP, either TEV-50717 or matching placebo. Maximal dose according to the titration scheme is presented.

Table 5: Drug Kits based on Titration Scheme –Impaired CYP2D6 Metabolizers

Weight Group	12kg to <17kg		17kg to <30kg		30kg to <40kg		≥40kg	
Visit	Dose (mg/day)	Tablets (mg)	Dose (mg/day)	Tablets (mg)	Dose (mg/day)	Tablets (mg)	Dose (mg/day)	Tablets (mg)
Day 2	6	6	6	6	6	6	6	6
Week 1	6	6	6	6	6	6	12	6,6
Week 2	6	6	12	6,6	12	6,6	18	9,9
Week 3	12	6,6	18	9,9	18	9,9	24	12,12
Week 4	12	6,6	18	9,9	24	12,12	30	15,15
Week 5	12	6,6	18	9,9	24	12,12	36	18,18
Week 6	12	6,6	18	9,9	24	12,12	36	9,9 AM 9,9 PM
Week 7	12	6,6	18	9,9	24	6,6 AM 6,6 PM	36	9,9 AM 9,9 PM

Dose and tablets refer to IMP, either TEV-50717 or matching placebo. Maximal dose according to the titration scheme is presented.

The number of tablets used is the number of tablets dispensed minus the number of tablets returned. If a patient does not return a pill bottle, it will be assumed that the patient took no IMP from that bottle for the purposes of calculating compliance.

The denominator of "expected number of tablets" will be calculated based on patients' participation in the study (ie, last day of study drug – first day of study drug + 1) and the

expected dose according to the RTSM system. Adjustments may be required for the doses split between the morning and the evening dose. Of note, the protocol allows the investigator not to titrate a patient up to the maximal possible dose by weight group and CYP2D6 impairment status; therefore the expected number of tablets will be calculated individually. Moreover, drug suspensions and reductions based on investigator instructions are considered as change in the expected number of tablets and are not considered non-compliance. For patients who complete the study, the last day of study drug used in the denominator calculation will be the last expected dosing day. For early termination patients, the last day of study drug used in the denominator calculation will be the last expected dosing day before termination day.

A patient will be deemed compliant over the treatment period if the patient has taken 80% to 105% of the expected tablets of study drug.

Treatment compliance (%) will be summarized as continuous data using descriptive statistics. In addition, treatment compliance will be summarized as categorical data using descriptive statistics for the following categories: <80%, 80% to 105%, >105%. Compliance will be presented by visit, by overall in the titration and maintenance periods, and by overall in the treatment period.

An additional assessment of the IMP compliance may be performed while taking into account the patient's dose.

5.9. Study Protocol Deviations

Data from patients with any important protocol deviations (as recorded in the Clinical Trial Management System) during the study will be summarized overall and for each category using descriptive statistics. All protocol deviations will be listed.

In addition, patients with any COVID-19 related protocol deviations will be summarized and listed separately.

Reasons for exclusion from the PP analysis set (as evaluated in the final blinded statistical data review meeting; see Section 3.4) will be summarized using descriptive statistics.

6. EFFICACY ANALYSIS

6.1. General

The mITT, mITTA, mITTB, ITT, and PP analysis sets (Section 3) will be used for all the efficacy analyses.

The mITT analysis set will be regarded as the main set for the analysis of the primary endpoint. It will also be used for all the efficacy endpoints unless otherwise specified. In addition, the mITTA, mITTB, ITT, and PP analysis sets will be used for the sensitivity analyses of the primary and key secondary endpoints.

Summaries will be presented by each scheduled visit (until Week 15) and treatment group, unless specified otherwise.

All p-values for treatment effect (TEV-50717 vs. placebo) will correspond to a 1-sided test for superiority.

6.2. Primary Efficacy Endpoint and Analysis

6.2.1. Primary Endpoint

6.2.2. Definition

The primary efficacy endpoint is the change from baseline to week 15 in the MD-CRS part II total score (movement disorder severity, centrally read) (TEV-50717 versus placebo).

6.2.3. Primary Efficacy Analysis

The primary analysis will be a mixed-model repeated-measures (MMRM) with the change in MD-CRS part II total score as the dependent variable. The model will include fixed effects for treatment group, week (as a categorical variable with 3 levels: weeks 9, 12, and 15), and treatment group by week interaction. The baseline MD-CRS part II total score, age group at baseline (2 levels: 6 to <12 years; 12 through 18 years, inclusive), and region (US; non-US) will be included as covariates. The unstructured covariance matrix will be used in the model. In cases where the model does not converge, the Maximum-Likelihood estimation method will be used instead of the default Restricted Maximum-Likelihood. If the model still does not converge, then a simpler covariance structure with fewer parameters will be used, according to the following order: Heterogeneous Autoregressive (1), Heterogeneous Compound Symmetry, Autoregressive (1), and Compound Symmetry. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom.

Missing data will be classified as MAR and MNAR, as described in Section 4.3. MNAR data will be imputed using the jump-to-reference method, and MAR data will be imputed based on the randomized treatment group. The details are described below:

1. Missing data will be imputed in 2 separate subgroups. The MAR subgroup will contain all patients randomized to a TEV-50717 treatment group without missing data at week 15 or have a MAR missing data mechanism. The MNAR subgroup will contain all patients randomized to placebo as well as patients randomized to receive TEV-50717 with missing data at week 15 with an MNAR missing data mechanism.
2. Data will be imputed for patients with missing data at week 15 using a predictive mean matching multiple imputation method. Multiple imputations will be performed within each subgroup:
    a. Under a MNAR data assumption, patients in the MNAR subgroup will have their missing data at week 15 imputed as if they behaved like placebo treated patients after dropping out. Only placebo patients with complete data will be included in the imputation model.
    b. All other patients will have data imputed from a model derived from their assigned treatment group (imputation under MAR assumption).

    The imputation models will include age group at baseline, region, baseline MD-CRS part II total score, and MD-CRS part II total scores at week 9 and week 12. After obtaining complete data sets for both subgroups, they will be combined and the complete data set will be used in the analysis.
3. The resulting complete, imputed datasets will each be analyzed using the analysis model described above and the resulting statistics combined using methodology presented by Rubin (1987) and Little and Rubin (2002).

The complete multiple imputation methodology, along with SAS code, is detailed in Appendix 1. Difference in least squares (LS) mean of the change in MD-CRS part II total score from baseline to week 15 (TEV-50717 versus placebo) will be compared using a 1-sided test for superiority at a nominal significance level of α=0.025.

The LS mean and SE for the treatment groups, the LS mean difference, 2-sided 95% confidence interval, and p-value for the comparison (TEV-50717 versus placebo) at week 15 will be presented.

6.2.4. Sensitivity Analysis

To assess the robustness of the primary efficacy analysis, sensitivity analyses will include the following:

Sensitivity analyses for assumptions of missing data mechanism:

- tipping point (MNAR)
- MAR multiple imputation

Sensitivity analysis for the statistical model:

- analysis of covariance (ANCOVA)
- primary model, including only the following covariates: treatment group, week, and treatment group by week interaction

Other:

- primary model repeated for the ITT analysis set;
- primary model repeated for the mITTA analysis set;
- primary model repeated for the mITTB analysis set;
- primary model repeated for the PP analysis set;
- ANCOVA models for MD-CRS part II total score change from baseline to week 9 and to week 12.

6.2.4.1. Sensitivity Analyses for Assumptions of Missing Data Mechanism

Sensitivity analyses for assumptions of missing data mechanism include the MNAR tipping point analysis and the MAR multiple imputations.

6.2.4.1.1. MNAR Tipping Point Analysis

Details for tipping point analysis are described below:

a. For patients that completed the study, intermittent missing data are considered as MAR and will be imputed by the predictive mean matching multiple imputation method (Heitjan and Little 1991, Schenker and Taylor 1996). The imputation model will include age group at baseline, region, treatment group and MD-CRS part II total score at baseline, week 9 and week 12.

b. For patients who discontinued due to any reason, all missing data will be imputed using the same predictive mean matching multiple imputation method. The imputed values will be adjusted under a MNAR assumption:

   – Patients randomized to TEV-50717: the first missing value after patients drop out will be analyzed assuming the treatment effect is worsened by δ (where $\delta_1 = 0$ to 2 or estimated treatment effect based on the primary analysis, whichever is higher, in step of 0.2) compared to the patients who have no missing value. $\delta \geq 0$ will be used as a shift to adjust the imputed values, since higher MD-CRS score corresponds to the more severe movement disorder.

   – Patients treated with placebo: the observed treatment effect for placebo-treated patients will be used for imputing missing value in the placebo population for all shift parameters.

   The shift will be applied only for the first visit with missing data and will not be applied for later time points as these values are shifted through the condition on the previous time points. For each δ value, 500 imputed datasets will be obtained.

c. The resulting complete, imputed datasets will each be analyzed using the same model as the primary analysis model, and the resulting statistics combined using methodology provided by Rubin (1987) and Little and Rubin (2002). Refer to Appendix 1; start at Step 2.

The complete multiple imputation methodology, along with the SAS code, is detailed in Appendix 2.

6.2.4.1.2. MAR Multiple Imputation

All missing data will be imputed using the predictive mean matching multiple imputation method (Heitjan and Little 1991, Schenker and Taylor 1996) under MAR data assumption using fully conditional specification (FCS) method. The imputation model will include age group at baseline, region, treatment arm, MD-CRS part II total score at baseline, week 9 and week 12.

The imputation under the MAR assumption corresponds to a full treatment effect observed in the mITT population. Therefore, the results from the tipping point analysis with $\delta = 0$ will be used for this sensitivity analysis; see Section 6.2.4.1.1 for details.

6.2.4.2. Sensitivity Analysis for the Statistical Model

Sensitivity analyses for the statistical model include the analysis of covariance (ANCOVA) and the primary model rerun with a reduced number of covariates.

6.2.4.2.1. Analysis of Covariance

The ANCOVA model will be used for week 15 MD-CRS part II total score change from baseline data following imputation using the same approach as described in Section 6.2.3, steps 1-2. The ANCOVA model will replace the MMRM analysis in step 3 of Section 6.2.3 and will include age group at baseline, region, baseline MD-CRS part II total score, and treatment as covariates. The SAS code for this model is as follows:

```
proc mixed;
  class arm agegrp region;
  model chg=arm agegrp region base;
   lsmeans arm / diff cl;
where avisitn=15;
run;
```

6.2.4.2.2. Primary Analysis With Reduced Number of Covariates

Primary analysis, including only the following covariates: treatment group, week, and treatment group by week interaction will be run as described in Section 6.2.3. The MAR/MNAR multiple imputation models will include only these limited covariates. After obtaining complete data sets for MAR/MNAR subgroups, they will be combined and the complete data set will be used in the analysis with the same covariates.

6.2.4.3. Other Sensitivity Analyses

Other sensitivity analyses will include the primary model repeated on the different subsets of patients and at the different time points.

The primary model as described in Section 6.2.3 will be repeated using

- ITT analysis set;
- mITTA analysis set;
- mITTB analysis set;
- PP analysis set.

ANCOVA model, as described in Section 6.2.4.2.1, but without the multiple imputation, will be applied for MD-CRS part II total score change from baseline to

- week 9;
- week 12;
- week 15.

6.2.5. Retrospective Power and Sample Size

The following analyses will be performed to quantify the impact of the reduction in sample size on the power, based on the actual final data:

- Conditional power for n=185 (the initially planned minimal sample size), based on the observed effect size at the end of the study
- Sample size required to power the study (90%) at α=0.025, based on the observed effect size at the end of the study
- Effect size and difference in change in MD-CRS II based on the observed SD needed in order to be powered (90%) for the given final sample size

The retrospective analyses will be conducted using the definitions and methods that were used in the interim analysis, see Section 10.4 and 10.5.

6.3. Secondary Efficacy Endpoints and Analysis

6.3.1. Key Secondary Efficacy Endpoints

The key secondary endpoints are the following:

1. change from baseline to week 15 in centrally read MD-CRS part I total score (TEV-50717 versus placebo)
2. CaGI-I at week 15 (TEV-50717 versus placebo)
3. CGI-I at week 15 (TEV-50717 versus placebo)
4. change from baseline to week 15 in centrally read UHDRS-TMC (TEV-50717 versus placebo)
5. change from baseline to week 15 in centrally read UHDRS-TMD (TEV-50717 versus placebo)

For a detailed description of scales, refer to the Section 6.2 of the study protocol.

6.3.1.1. Key Secondary Efficacy Endpoints Analysis

Analyses of the key secondary endpoints will be performed in a hierarchical manner as follows:

1. MD-CRS part I total score change from baseline to week 15 will be analyzed using a MMRM in the same fashion as the primary analysis, with the exception that the baseline value of MD-CRS part I total score will be included as the covariate instead of MD-CRS part II total score.

2. CaGI-I at week 15 will be analyzed using an MMRM model in the same fashion as the primary analysis, with the exceptions that the baseline value of CGI-S will be included as the covariate instead of MD-CRS part II total score and that the week variable will include 5 levels (weeks 3, 7, 9, 12, and 15).
3. CGI-I at week 15 will be analyzed using an MMRM model in the same fashion as the primary analysis, with the exceptions that the baseline value of CGI-S will be included as the covariate instead of MD-CRS part II total score and that the week variable will include 5 levels (weeks 3, 7, 9, 12, and 15).
4. UHDRS-TMC (centrally-read) change from baseline to week 15 will be analyzed using an MMRM model in the same fashion as the primary analysis, with the exception that the baseline value of UHDRS-TMC (centrally-read) will be included as the covariate instead of MD-CRS part II total score.
5. UHDRS-TMD (centrally-read) change from baseline to week 15 will be analyzed using an MMRM model in the same fashion as the primary analysis, with the exception that the baseline value of UHDRS-TMD (centrally-read) will be included as the covariate instead of MD-CRS part II total score.

For CaGI-I and CGI-I, in Appendix 1, Step 1 should begin with week 3, week 7 and only then continue to week 9.

See Section 6.5.3 for definition of UHDRS-TMC and TMD, and Section 7 for description of multiplicity control of the key secondary endpoints.

6.3.1.2. Key Secondary Efficacy Endpoints Sensitivity Analysis

Sensitivity analyses for the key secondary endpoints will be performed in the same fashion as the primary endpoint. See Section 6.2.4 for details.

Some specific notes:

- The corresponding baseline or post-baseline assessment will be used for each secondary endpoint.
- Tipping point (MNAR) – For CaGI-I and CGI-I, the imputation model will include age group at baseline, region, treatment group, and CGI-S core at baseline, as well as the CaGI-I or CGI-I score at week 3, 7, 9 and week 12. MAR multiple imputation will be analyzed in a similar fashion with the same covariates.
- ANCOVA models will be fitted for change from baseline in MD-CRS part I total score to weeks 9, 12 and 15, and CaGI-I and CGI-I scores at weeks 3, 7, 9, 12 and 15.

6.4. Subgroup Analyses

Analyses of the primary and key secondary endpoints by subgroup below will be conducted using the mITT population:

- Gender: male and female;
- Race Group: white, black or African American, and other;

- Baseline Age Group (6-<12 years, 12-18 years inclusive);
- CYP2D6 impairment status (Yes/No);
- US vs outside of US (non-US).

As the total sample size is approximately 65 patients, the sub-group analysis will include descriptive statistics by subgroup, and no statistical models will be fitted. The subgroup analyses will be performed provided the minimum size of the smaller subgroup is at least 5 patients. Otherwise, the main analysis will be repeated on the larger subgroup.

6.5. Other Efficacy Endpoints and Analysis

Other efficacy endpoints (change from baseline to week 15/proportion at week 15; TEV-50717 versus placebo) are the following:

- MD-CRS physician rated:
  - MD-CRS part I total score change from baseline to week 15
  - MD-CRS part II total score change from baseline to week 15
  - MD-CRS Global Index (calculated from MD CRS parts I and II total scores as assessed by physician)
- MD-CRS Global Index (calculated from MD-CRS parts I and II total scores, centrally read)
- UHDRS:
  - UHDRS-TMS (physician rated)
  - UHDRS-TMC (physician rated)
  - UHDRS-TMD (physician rated)
- PEDI CAT
- PedsQL
- PGI-I
- CGI-S
- Response Analysis
  - CaGI-I response, defined as patients who are described by the caregiver as “Much Improved” or “Very Much Improved” in the CaGI-I score
  - CGI-I response, defined as patients who are described as “Much Improved” or “Very Much Improved” in the CGI-I score
  - CGI-S response, defined as patients who have a reduction of ≥1 point in the CGI-S score
  - PGI-I response, defined as patients who are described as “Much Improved” or “Somewhat Improved” in the PGI-I score

The mITT analysis set will be used for the analysis.

6.5.1. MD-CRS Global Index (calculated from MD-CRS parts I and II total scores, centrally read)

6.5.1.1. Definition

The MD-CRS Global Index is a global measure of the MD-CRS that consolidates the information from parts I and II using the method of weighted means of the 2 normalized indexes obtained from each part. The minimum score is 0 and the maximum score is 1.

The standardized/normalized score for each item of MD-CRS parts I and II with value X is calculated using the formula

$$X_{st} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

where X_{max} is the maximum value for the score, and X_{min} is the minimum value for the score, or 4 and 0 respectively.

The normalized index for the scale, MD-CRS parts I or II, Index I or II, is calculated as the mean value of X_{st}.

The MD-CRS Global Index is the weighted mean of the normalized indexes for centrally read MD-CRS parts I and II using the formula:

$$Global\ Index = \frac{n_I \times Index_I + n_{II} \times Index_{II}}{n_I + n_{II}}$$

where n_I and n_{II} are the numbers of items in MD-CRS parts I and II respectively.

6.5.1.2. Analysis

The MD-CRS Global Index change from baseline to week 15 data will be analyzed using MMRM. The MMRM model will include age group at baseline, week (3 levels: weeks 9, 12, and 15), region, corresponding baseline value and treatment as fixed effects. The SAS code is below:

```
proc mixed Method=REML;
  class usubjid arm avisitn agegrp region;
  model chg=arm | avisitn agegrp region base / ddfm=kr;
  repeated avisitn / type=un subject=usubjid;
  lsmeans arm*avisitn / diff cl;
  ods output diffs=diff(where=(avisitn=_avisitn) lsmeans=lsm;
Run;
```

The LS mean and SE for the treatment groups, the LS mean difference, 2-sided 95% confidence interval (CI), and 1-sided p-value for the comparison (TEV-50717 versus placebo) at Week 15 will be presented. The p-value will be considered to be a nominal p-value.

A descriptive summary will be provided for all assessments and change from baseline, including the follow-up period.

6.5.2. MD-CRS Physician Rated

6.5.2.1. Definition

The local, physician rated MD-CRS part I and MD-CRS part II are additional efficacy endpoints of the study. The physician rated MD-CRS part I and MD-CRS part II endpoints include:

- MD-CRS part I total score (general assessment, physician rated) change from baseline to week 15
- MD-CRS part II total score (general assessment, physician rated) change from baseline to week 15
- MD-CRS Global Index (calculated from MD CRS parts I and II total scores as assessed by physician) change from baseline to Week 15

The calculation as described in Section 6.5.1.1 will be performed for physician-rated MD-CRS assessments to yield the MD-CRS Global Index.

6.5.2.2. Analysis

The data will be analyzed using MMRM similarly to the description in Section 6.5.1.2.

6.5.3. Unified Huntington's Disease Rating Scale

6.5.3.1. Definition

UHDRS-TMS:

The UHDRS-TMS (physician rated) comprises a broad assessment of features associated with HD. The TMS component of UHDRS comprises 31 assessments from the 15 items of the UHDRS. The TMS is calculated as the sum of the 31 motor assessments, each of which range between 0 to 4. The minimum score is 0 (absent) and the maximum score is 124 (worst).

The UHDRS-TMS is administered at baseline and weeks 9, 12, and 15/ET visits.

UHDRS-TMC:

The UHDRS-TMC is part of the UHDRS-TMS assessment and evaluates the severity of chorea in the face, mouth, trunk, and the 4 extremities. The minimum score is 0 (absent) and the maximum score is 28 (marked/prolonged).

UHDRS-TMD:

The UHDRS-TMD is part of the UHDRS-TMS assessment and assesses the severity of dystonia in the trunk and the 4 extremities. The minimum score is 0 (absent) and the maximum score is 20 (marked/prolonged).

If responses of up to 25% of items are missing in UHDRS-TMS, UHDRS-TMC, or UHDRS-TMD ("scale"), the missing responses will be replaced by the average of the remaining responses

within each of the scales. If 8 or more of the items are missing, the missing items will not be replaced and the scale will be set to missing.

The UHDRS-TMC and TMD are administered at baseline and at weeks 9, 12, and 15/ET visits.

6.5.3.2. Analysis

The UHDRS-TMS, UHDRS-TMC (physician rated) and UHDRS-TMD (physician rated) change from baseline to week 15 data will be analyzed using MMRM similarly to the description in Section 6.5.1.2.

6.5.4. Pediatric Evaluation of Disability Inventory Computer Adaptive Test Definition

The PEDI-CAT is a clinical assessment for children and youth. The PEDI-CAT (activities of daily living [ADL], parent/caregiver completed, content balanced version) measures function in 4 domains: (1) Daily Activities; (2) Mobility; (3) Social/Cognitive, and (4) Responsibility. Each domain is self-contained and can be used separately or with other domains. The content balanced version presents a balance of items from each of the Daily Activities domain's content areas (Getting Dressed, Keeping Clean, Home Tasks, and Eating & Mealtime). A total of approximately 30 items are administered (total number of items administered is dependent upon the responses and decision tree path). The PEDI-CAT software utilizes Item Response Theory statistical models to estimate a child's abilities from a minimal number of the most relevant items or from a set number of items within each domain. All respondents begin with the same item in each domain in the middle of the range of difficulty or responsibility and the response to that item then dictates which item will appear next (a harder or easier item), thus tailoring the items to the child and avoiding irrelevant items. The CAT program then displays the results: normative standard scores, scaled scores and the SE. The scales are available for 21 age groups as intervals of 1 year.

Normative (fit) scores are based on a child's chronological age and intended for use by clinicians so that they may interpret a particular child's functioning relative to others of the same age. This metric is of less relevance for the studied population and therefore will not be assessed.

Scaled scores provide a way to look at a child's current functional skills and progress in these skills over time. Scaled scores are especially helpful in documenting improvements in functional skills for children not expected to exhibit or regain normative levels of functioning. The scaled scores are based on an estimate of the placement of an individual child along the hierarchical scale within each domain. The PEDI-CAT scaled scores are currently on a 20 to 80 scale metric. See PEDI-CAT (Version 1.3.6) Development, Standardization and Administration Manual.

The SE scores indicate the level of the test precision, and are expected to be up to 0.5.

The PEDI-CAT (ADL, parent/caregiver completed, content balanced version) is administered at baseline and at week 15/ET visits.

6.5.4.1. PED-CAT Analysis

The PEDI-CAT scaled score change from baseline to week 15 data will be analyzed using an ANCOVA model with age group at baseline, region, baseline score, and treatment as fixed effects based on the mITT analysis set. The ANCOVA model will be applied as in Section 6.2.4.2.1.

A descriptive summary of the PEDI-CAT scaled score will be provided for all measurements and change from baseline.

6.5.5. The CP module of the PedsQL (QoL, patient/caregiver)

6.5.5.1. Definition

The Pediatric Quality of Life Inventory (PedsQL™) is a health-related quality-of-life instrument that consists of a well-validated generic core measure and some condition and disease-specific modules (Varni et al 2006). The 35-item PedsQL 3.0 CP module encompasses 7 scales: (1) Daily Activities (9 items); (2) School Activities (4 items); (3) Movement and Balance (5 items); (4) Pain and Hurt (4 items); (5) Fatigue (4 items); (6) Eating Activities (5 items); and (7) Speech and Communication (4 items).

The scales comprise parallel child self-report and parent proxy-report formats. Child self-report includes ages 5 to 7 years, 8 to 12 years, and 13 to 18 years. A parent proxy report is included in reports for ages 2 to 4 years (toddler), 5 to 7 years (young child), 8 to 12 years (child), and 13 to 18 years (adolescent, not administered in the study), and assesses parents' perceptions of their child's health-related quality of life (QoL). The instructions ask how much of a problem each item has been during the past 1 month. A 5-point response scale is utilized across child self-report and parent proxy report as follows:

0=never a problem;

1=almost never a problem;

2=sometimes a problem;

3=often a problem;

4=almost always a problem.

To further increase the ease of use for the young child self-report (ages 5–7 years), the response scale is reworded and simplified to a 3-point scale (0=not at all a problem; 2=sometimes a problem; 4=a lot of a problem).

Items are reverse scored and linearly transformed to a 0–100 scale (0=100, 1=75, 2=50, 3=25, 4=0), so that higher scores indicate better quality of life, i.e. fewer symptoms or problems. Scale Scores are computed as the sum of the items divided by the number of items answered (this accounts for missing data). If more than 50% of the items in the scale are missing, the Scale Score is not computed. This computation is consistent with the previous PedsQL peer-reviewed publications as well as other well-established Health-related quality of life measures.

A total score will be computed as the sum of the 35 items divided by the number of items answered and will not be computed if more than 50% of the items are missing. Each teen/child self-report and parent proxy-report will be assessed separately.

The PedsQL is administered at baseline and at week 15/ET visits.

6.5.5.2. Analysis

The PedsQL total score per each teen/child self-report /parent proxy-report and change from baseline to week 15 will be analyzed using an ANCOVA model with region, baseline score, and

treatment as fixed effects based on the mITT analysis set. The ANCOVA model will be applied as in Section 6.2.4.2.1.

A descriptive summary of each PedsQL teen/child self-report /parent proxy-report total and scale scores and change from baseline will be provided.

6.5.6. PGI-I Scale (global, patient/caregiver)

6.5.6.1. Definition

For a detailed description of scales, refer to the Section 6.3.7 of the study protocol.

The PGI-I is administered at week 3, week 7, week 9, week 12, week 15/ET, and week 16/EOS visits. Input from the caregiver/adult is permitted.

6.5.6.2. Analysis

The PGI-I at week 15 data will be analyzed using MMRM based on the mITT analysis set similarly to the description in Section 6.5.1.2 (except that the value at week 15 will be analyzed rather than the change from baseline). The MMRM model will include age group at baseline, week (3 levels: weeks 9, 12, and 15), region, and treatment as fixed effects. No baseline assessment will be used in the model.

6.5.7. CGI-S Scale (global, physician rated)

6.5.7.1. Definition

For a detailed description of the scale, refer to the Section 6.3.8 of the study protocol.

The CGI-S is administered at screening, baseline, week 3, week 7, week 9, week 12, week 15/ET, and week 16/EOS visits.

6.5.7.2. Analysis

The CGI-S change from baseline to week 15 data will be analyzed using MMRM similarly to the description in Section 6.5.1.2. CGI-S baseline assessment will be used as the corresponding baseline value.

6.5.8. Response Analysis

The following response analyses will be performed:

- CaGI-I response, defined as patients who are described by the caregiver as “Much Improved” or “Very Much Improved” in the CaGI-I score;
- CGI-I response, defined as patients who are described as “Much Improved” or “Very Much Improved” in the CGI-I score;
- CGI-S response, defined as patients who have a reduction of ≥1 point in the CGI-S score;
- PGI-I response, defined as patients who are described as “Much Improved” or “Somewhat Improved” in the PGI-I score.

The proportion of patients answering each response definition above will be calculated as the number of patients answering each response definition above in each treatment group divided by the number of patients in the given treatment group at each visit where the corresponding scale was assessed. The proportion of responder patients will be compared between groups using the Cochran–Mantel–Haenszel test adjusting for age group at baseline and region.

The responder patient counts and percentages will be presented at each visit.

As a sensitivity analysis, proportion of responder patients for each criterion will also be summarized and analyzed as described above by classifying patients with missing data at a visit as a non-responder, meaning the entire mITT analysis set population participates in the analysis.

6.6. Minimal Clinically Important Difference (MCID) and Miminal Clinically Important Change (MCIC) Exploratory Analysis

6.6.1. Descriptive Statistics of MD-CRS part II vs Global Impression of Improvement at Week 15

For each anchor, the change from baseline in MD-CRS part II vs CaGI-I, CGI-I, and PGI-I at week 15 will be presented graphically.

In addition, the mean change from baseline in MD-CRS part II vs global impression of improvement at week 15 will be summarized by treatment group as follows:

For CaGI-I, CGI-I as anchors:

	Mean Change from Baseline in MD-CRS part II					
	TEV-50717		Placebo		All	
	N	Mean (SE)	N	Mean (SE)	N	Mean (SE)
Very Much Improved						
Much Improved						
Minimally Improved						
Not Changed						
Minimally Worse						
Much Worse						

For PGI-I as anchor:

	Mean Change from Baseline in MD-CRS part II					
	TEV-50717		Placebo		All	
	N	Mean (SE)	N	Mean (SE)	N	Mean (SE)
Much Improved						
Somewhat Improved						
Not Changed						
Somewhat Worse						
Much Worse						

6.6.2. MCID Analysis

The MCID analysis will be based on the method described by Hauser et al (2014). According to this method, the MCID is defined as the difference between patients in the TEV-50717 group who were rated “Minimally Improved” and patients in the placebo group who were rated “No Change”. The calculation will be performed for each anchor (CaGI-I, CGI-I and PGI-I) separately[2].

The MCID analyses will be performed if there are at least 5 patients in the TEV-50717 group and at least 5 patients in the placebo group who contribute to the calculation.

[2] For PGI-I, the level “Somewhat Improved” will be used instead of “Minimally Improved”.

6.6.3. MCIC Analysis

The MCIC analysis will be based on the method described by Hauser et al (2022). According to this method, the MCIC is defined as the mean change from baseline in MD-CRS part II in patients treated with TEV-50717 who were rated "Minimally Improved".

If the mean change from baseline in MD-CRS part II at week 15 for the patients in the placebo group who were rated "No Change" is negative, indicating a placebo effect in this subgroup, then the MCIC analysis will not be carried out.

The MCIC analyses will be performed if there are at least 5 patients in the TEV-50717 group who contribute to the calculation.

6.6.4. Descriptive Statistics of Improvement in MD-CRS part II

The following descriptive statistics for the change from baseline in MD-CRS part II by treatment group will be added: a histogram of the change; cumulative distribution of improvement (i.e., proportion of patients with change ≤ 0; change ≤ -1 ; change ≤ -2; change ≤ -3, change ≤ -4, …).

7. MULTIPLE COMPARISONS AND MULTIPLICITY

The primary efficacy endpoint will be tested at the 1-sided significance level of α=0.025.

If the primary endpoint is statistically significant (p-value ≤0.025), the 5 key secondary hypotheses will be tested using a hierarchical approach at the 1-sided significance level of α=0.025, in the following order: (1) MD-CRS part I total score, (2) CaGI-I, (3) CGI-I, (4) UHDRS-TMC (centrally-read), and (5) UHDRS-TMD (centrally-read)

[redacted]

No multiplicity control will be applied to the sensitivity analysis or other endpoints.

8. SAFETY ANALYSIS

8.1. General

Safety analyses will be performed on the safety analysis set (Section 3.3), unless otherwise noted. Summaries will be presented by treatment group and for all patients, unless otherwise stated. Selected safety data might also be presented by age group, as applicable.

For continuous variables, descriptive statistics will be provided for actual values and changes from baseline to each time point. For categorical variables, patient counts and percentages will be provided.

Safety assessments and time points are provided in Table 3 of the study protocol.

8.2. Duration of Exposure to Study Drug

Duration of exposure to study drug (days) for individual patients is the number of days patient received drug (last day of study drug – first day of study drug + 1). Duration of treatment (days) will be summarized using descriptive statistics. Weeks on treatment using the categories ≤3 weeks, >3 to ≤7 weeks, >7 to ≤9 weeks, >9 to ≤12 weeks, >12 to ≤15 and >15 weeks will also be summarized using descriptive statistics.

8.3. Adverse Events

Adverse events will be collected and recorded from the time a patient signs the informed consent to the end of follow-up period. For this study, the follow-up period for recording of adverse events is defined as 2 weeks after the last dose of IMP or up to baseline/day 1 in the open-label safety extension study TV50717-CNS-30081 (whichever comes first) or 2 weeks after last site visit for patients who terminated early from the study.

All adverse events will be coded using MedDRA. Each patient will be counted only once in each preferred term or SOC category for the analyses of safety. Unless otherwise specified, adverse events will be summarized by SOC and preferred term, with SOCs and preferred terms within SOCs presented in descending order of patient incidence. Summaries will be presented for all adverse events (overall and by severity), adverse events determined by the investigator to be related to test IMP (ie, reasonable possibility; see Table 7 of the study protocol - defined as related or with missing relationship), overall and by severity, serious adverse events, and adverse events causing withdrawal from the study; Most common adverse events, defined as occurring in > 4% of the patients in either of the treatment groups overall, and by preferred term. Summaries will be presented by treatment group and for all patients.

Summaries will include treatment-emergent adverse events which are defined as adverse events occurring at or after the first dose of the IMP. The listings will include all adverse events recorded.

Patient listings of all adverse events, serious adverse events, adverse events leading to study drug discontinuation, adverse events leading to dose interruption and adverse events leading to death will be presented. Patient’s dose at the time of the AE onset will be included in all listings.

8.4. Deaths

If any patient dies during the study, a listing of deaths will be provided, and all relevant information will be discussed in the patient narrative included in the CSR.

8.5. Clinical Laboratory Tests

Summary statistics for chemistry, hematology and coagulation, and urinalysis laboratory tests are assessed at screening and week 15. The assessment will be repeated at week 16 visit for patients with clinically significant laboratory abnormalities at week 15. Laboratory values and changes from baseline to week 15 and the last assessment as defined in Section 4.4 will be presented.

A list of laboratory tests is included in Table 8 of the study protocol.

Shifts (below, within, and above the normal range) from baseline to last assessment will be summarized using patient counts for all the laboratory tests as applicable. Only patients with both baseline and post-baseline assessments will be summarized in the shift tables.

Only central laboratory assessments will be summarized; local laboratory assessments will be listed. Values below the lower limit of quantification or above the upper limit of quantification will be assessed on a case-by-case basis if present in the data and discussed in the final blinded SDR meeting.

8.6. Physical Examinations

Physical examinations will be performed at the time points detailed in Table 3 of the study protocol. Physical examinations will not be summarized. Any finding will be listed.

Weight, height and BMI will be assessed as a part of the vital signs (Section 8.7).

8.7. Vital Signs

Vital signs (blood pressure [BP; systolic/diastolic], respiratory rate, body temperature, and pulse) will be measured at the time points detailed in Table 3 of the study protocol.

Vital signs (supine or semi erect/seated) values and changes from baseline to each visit will be summarized using descriptive statistics. The last assessment will be defined as in Section 4.4.

Orthostatic systolic and diastolic BP and pulse will be calculated as supine or semi erect/seated measurement minus standing measurement and values and changes from baseline to each visit will be summarized using descriptive statistics. Any cases where standing measurements were taken less than 3 minutes after the supine or semi erect/seated measurement will be excluded from the analysis.

Orthostatic hypotension (determined by BP measurements only; positive value in case of abnormality) is defined as having either a ≥20 mmHg reduction from supine to standing position in systolic blood pressure (SBP) or ≥10-mmHg reduction from supine to standing position in diastolic blood pressure (DBP) or both. Orthostatic tachycardia is defined as pulse increase ≥20 beats per minute (bpm) from supine to standing position.

Summaries of potentially clinically significant abnormal values will include all post-baseline values (including scheduled, unscheduled, and withdrawal visits). The incidence of potentially

clinically significant abnormal values will be summarized using descriptive statistics with the criteria specified in Table 6.

Table 6 specifies the criteria for identifying vital signs as potentially clinically significant abnormal values.

Table 6: Criteria for Potentially Clinically Significant Vital Signs

Vital Sign (unit)	Age	Potentially Clinically Significant Low Criterion value	Potentially Clinically Significant High Criterion value
SBP (supine/ seated/ semi-erect) (mmHg)	6-12	Value ≤70 and≥20 decrease from baseline	Value ≥120 and ≥20 increase from baseline
	13-18	Value ≤ 90 and ≥20 decrease from baseline	Value ≥135 and ≥20 increase from baseline
DBP (supine/ seated/ semi-erect) (mmHg)	6-12	Value ≤40 and ≥15 decrease from baseline	Value ≥80 and ≥15 increase from baseline
	13-18	Value ≤50 and ≥15 decrease from baseline	Value ≥90 and ≥15 increase from baseline
Pulse rate (supine/ seated/ semi-erect) (bpm)	6-10	Value ≤60 and ≥15 decrease from baseline	Value ≥135 and ≥15 increase from baseline
	11-18	Value ≤50 and ≥15 decrease from baseline	Value ≥120 and ≥15 increase from baseline
SBP orthostatic criteria (mmHg)	~	≥20 decrease from supine to standing position	NA
DBP orthostatic criteria (mmHg)	~	≥10 decrease from supine to standing position	NA
Pulse rate orthostatic criteria (bpm)	~	NA	≥20 increase from supine to standing position
Temperature (˚C)	~	NA	Value ≥38.3°C and ≥0.8°C increase from baseline

The BMI will be computed similarly to description in Section 5.3 at post-baseline visits using weight at visit and last available height measurement.

In addition, normal age and sex-based z-scores and percentiles for BMI will be determined using the WHO growth charts. Age and sex-based BMI categories include: Underweight (< 5 percentile), Normal (≥ 5 - < 85 percentile), Overweight (≥ 85 - < 95 percentile), Obese (≥ 95 percentile).

Descriptive statistics for weight, height, BMI, BMI World Health Organization adjusted z-scores and percentile will be provided. The categorical variables of BMI categories will be tabulated for each category as mentioned above. Data will be listed.

8.8. Electrocardiography

A 12-lead ECG will be recorded at the time points detailed in Table 3 of the study protocol. All ECGs will be performed after at least 5 minutes rest in a supine or semi-supine position. A qualified physician at a central diagnostic center will be interpreting the ECG. A central ECG Standard 12-lead ECG will record heart rate (HR), PR interval (Time from the beginning of the P wave to the beginning of the QRS complex in electrocardiogram), RR interval (Time between the start of one R wave and the start of the next R wave in the ECG), QT interval, Fridericia's corrected QT interval (QTcF), and QRS duration (Graphical deflections corresponding to ventricular depolarization in a typical electrocardiogram).

All ECG results outside of the reference ranges will be judged by the investigator as belonging to one of the following categories:

- abnormal and not clinically significant
- abnormal and clinically significant

If an abnormal post-baseline QTcF value >500 msec or change from screening >60 msec is found, the investigator must further evaluate QTcF based on protocol section 7.8.

Shifts (normal and abnormal) from baseline to overall result interpretation, each visit and last assessment, will be summarized using patient counts for the centrally-read interpretation. For overall result interpretation the worst post-baseline finding for the patient (the abnormal finding if there are both normal and abnormal findings) will be used in the summaries. Summary statistics for ECG variables values will be presented. Actual values and changes from baseline to each scheduled visit and last assessment will be summarized using descriptive statistics.

QTcF values will be classified as having QTc prolongation if any of the following conditions are met.

- Confirmed QTcF >450 msec
- Confirmed QTcF>480 msec
- Confirmed QTcF>500 msec
- Increase from screening QTcF > 60 msec
- Increase from screening confirmed QTcF > 60 msec

Screening is defined as the last non-missing ECG value prior to first dose IMP and the confirmed QTcF is defined as the average of up to 3 standard ECG values at a visit, or a single summary ECG. If there are more than 3 standard or more than one summary ECG at a visit, then use the last ECGs for the confirmed QTcF.

The number and percentage of patients with QTc prolongation will be summarized by treatment and overall, and by visit and last assessment.

The incidence of potentially clinically significant abnormal values for ECG variables will be summarized using descriptive statistics with the criteria presented in Table 7. The summary will be completed by treatment and overall, and age groups (6 to < 8 years, 8 to < 12 years, 12 to < 16 years, and ≥16 years), and overall using the age at screening.

Table 7: Potentially Clinically Significant Values for ECG Variables by Age

ECG parameter (unit) [a]	Age (years old)	Potentially Clinically Significant Low	Potentially Clinically Significant High
HR (bpm)	6 to <8	< 65	> 115
	8 to <12	< 55	> 110
	12 to <16	< 50	> 105
	≥16	< 50	> 100
PR interval (msec)	6 to <8	--	> 160
	8 to <12	--	> 175
	12 to <16	--	> 180
	≥16	--	> 200
QRS interval (msec)	6 to <8	--	> 100
	8 to <12	--	> 105
	12 to <16	--	> 110
	≥16	--	> 120

[a] bpm=beats per minute; HR=heart rate; PR interval =Time from the beginning of the P wave to the beginning of the QRS complex in electrocardiogram; RR interval =Time between the start of one R wave and the start of the next R wave in the ECG; QRS duration =Graphical deflections corresponding to ventricular depolarization in a typical electrocardiogram

8.9. Concomitant Medications or Therapies

Concomitant therapies and medications, including medications that are taken on an as needed basis and occasional therapies, will be monitored during the study. Details of prohibited medications may be found in Section 5.7 of the study protocol. All concomitant medications will be coded using the WHO Drug.

The incidence of concomitant therapies and medications will be summarized using descriptive statistics by therapeutic class category, preferred base name and preferred term. Patients are counted only once in each therapeutic class, once in each preferred base name, and only once in each preferred term category. Concomitant therapies and medications will include all medications taken after the first administration of the IMP up to the end of study as defined in the study protocol.

In order to determine whether a medication is concomitant, partial start dates with missing day and/or month will be imputed with the first day of the month/first month of the year, respectively and partial end dates with missing day and/or month will be imputed with the 28th day of the month/last month of the year respectively. Imputed dates will be used for calculations only, actual dates will be listed.

Prohibited medications that are associated with QTc prolongation are listed in Appendix H, Table 11, of the study protocol. Any incidence of these prohibited medications will be listed separately.

8.10. Children's Columbia-Suicide Severity Rating Scale

The frequency and severity of suicidal ideation or behavior according to the children's C-SSRS questionnaire will be presented for all patients aged ≥12 years by visit and by treatment group. A shift table to examine changes in children's C-SSRS categories from baseline (using the lifetime version of the questions), compared to the worst (highest) category during the treatment period will be presented by treatment and overall.

Patients will be placed into categories for suicidal ideation and suicidal behavior based on their responses to various questions. The suicidal ideation categories will be determined by the following 5 questions.

Table 8: Suicidal Ideation Category

Type	Category
Suicidal Ideation	(0) None – if response is No to Questions 1 and 2 (1) Wish to be Dead – if response to Question 1 is Yes and responses to Question 2 is No. (2) Non-Specific Active Suicidal Thoughts – if response to Question 2 is Yes and response to Questions 3-5 are No. (3) Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act – if response to question 3 is Yes and response to questions 4 and 5 are No. (4) Active Suicidal Ideation with Some Intent to Act, without Specific Plan – if response to Question 4 is Yes and response to Question 5 is No. (5) Active Suicidal Ideation with Specific Plan and Intent – if response to Question 5 is Yes.

For suicidal behavior, the following categories will be used. The categories will be determined based on the response to the questions.

Table 9: Suicidal Behavior Category

Type	Category
Suicidal Behavior	(6) Preparatory Acts or Behavior – if response to Preparatory Acts and Behavior is Yes and responses to Actual Attempt, Interrupted Attempt, Aborted Attempt, and Completed Suicide are No (7) Aborted Attempt – if response to Aborted Attempt is Yes and responses to Actual Attempt, Interrupted Attempt, and Completed Suicide are No. (8) Interrupted Attempt – if response to Interrupted Attempt is Yes and response to Actual Attempt, and Completed Suicide are No. (9) Actual Attempt – if response to Actual Attempt is Yes and Completed Suicide is No. (10) Completed Suicide - if response to Completed Suicide is Yes (0) None – if responses to all the above questions are No.

Suicidal ideation or behavior will be derived as the highest suicidal ideation or behavior score at the visit. Score of 0 represent "No Suicidal Ideation/Behavior".

The frequency and percentage of suicidal ideation, suicidal behavior, suicidal ideation or behavior, and self-injurious behavior without suicidal intent will be summarized by treatment and overall, and visit.

8.11. Child Behavior Checklist for Ages 6-18 (CBCL)

The CBCL assesses behavioral and emotional status in children ages 6 through 18 years of age as reported by the parent/caregiver (Achenbach and Ruffle 2000; Achenbach 2005). The full CBCL has 2 parts, a Competence Scale (Parts I to VII) and a Syndrome Scale (behavioral items).

8.11.1. Competence Scale

The Competence Scale (Parts I to VII) assesses various activities**,** interpersonal relationships, and academic performance. Competence assessment will be performed at screening and week 15/ET.

The competence scale is divided into 3 subscales – see Table 10.

Table 10: CBCL Competence Scale

Activities	**Social**	**School**
I.A. number of sports	III.A. number of organizations	VII.1. Mean performance
I.B. Mean part/skills	III.B. Mean Participation	VII.2. Special class
II.A. number of activities	V.1 number of friends	VII.3. Repeated grade
II.B. Mean part/skills	V.2 Frequency of contact	VII.4. School problems
IV.A. number of jobs	VI.A. Behavior with others	
IV.B. Mean job quality	VI.B. Behavior alone	

Total score and change from baseline for each subscale separately and a total score and change from baseline for all items will be summarized for each time point using descriptive statistics for continuous variables. Total scores are rounded to nearest 0.5.

Scores per item are calculated as follows in Table 11:

Table 11: CBCL Competence Scale Scoring

Subscale	Item	Scoring
Activities	I.A. number of sports	count the number of sports
	I.B. Mean part/skills	mean of up to 6 rates, 2 for each sport activity. Where 'below average' or 'Less than average'=0, 'Average'=1, 'above average' or 'More than average'=2
	II.A. number of activities	count the number of activities.
	II.B. Mean part/skills	mean of up to 6 rates, 2 for each activity. Where 'below average' or 'Less than average'=0, 'Average'=1, 'above average' or 'More than average'=2
	IV.A. number of jobs	count the number of jobs
	IV.B. Mean job quality	mean of up to 3 rates, 1 for each job. Where: 'below average'=0, 'Average'=1, 'above average''=2.
Social	III.A. number of organizations	count the number of organizations.
	III.B. Mean Participation	mean of up to 3 rates, 1 for each organization. Where: 'Less active'=0, 'Average'=1, 'More active'=2.
	V.1 number of friends	None'=0, '1'=1, '2 or 3' =2, '4 or more'=3.
	V.2 Frequency of contact	'Less than 1'=0, '1 or 2'=1, '3 or more'=2.
	VI.A. Behavior with others	mean of items a-c. Where: 'Worse'=0, 'Average'=1, 'Better'=2.
	VI.B. Behavior alone	rate for item d. Where: 'Worse'=0, 'Average'=1, 'Better'=2.
School	VII.1. Mean performance	Mean of up to 7 performance items. Where 'failing'=0, 'below average'=1, 'average'=2,'above average'=3.
	VII.2. Special class	'Yes='0', 'No='1'.
	VII.3. Repeated grade	'Yes='0', 'No='1'.
	VII.4. School problems	'Yes='0', 'No='1'.

Values of 'Do not know', 'Not applicable', or 'Has no brothers or sisters' will be ignored for mean calculations and logically skipped items will be imputed as 0 or none.

A subscale score is calculated as the sum of the item scores and the total score is calculated as the sum of the subscale scores. If one or more subscale scores are missing, the total score is not calculated.

If 1 item from activities and social subscales has a missing value, the mean of the other items of that subscale is substituted for that item. If more than 1 item is missing from either of these scales or any item is missing from the school scale, the respective scale is not to be scored.

Results will be listed.

8.11.2. Syndrome Scale

The Syndrome Scale comprises 113 questions related to problem behaviors. Syndrome assessment will be performed at screening, baseline, weeks 3, 7, 9, 12, 15/ET, 16/EOS.

For each item, the parent/caregiver will circle “0” if the item is not true of their child, “1” if the item is somewhat or sometimes true, and “2” if the item is very true or often true.

The Syndrome scale is divided into 8 subscales – see Table 12.

Table 12: CBCL Syndrome Scale

Anxious / Depressed		**Withdrawn / Depressed**	**Somatic Complaints**	**Social Problems**	**Thought Problems**	
14.Cries 30.FearsSchool 32.Perfect 35.Worthless 50.Fearful 71.SelfConsc 112.Worries	29.Fears 31.FearDoBad 33.Unloved 45.Nervous 52.Guilty 91.ThinksSuicide	5.EnjoysLittle 42.RatherBeAlone 65.Won'tTalk 69.Secretive 75.Shy 102.LacksEnergy 103.Sad 111.Withdrawn	47.Nightmares 49.Constipated 51.Dizzy 54.Tired 56a-g.Physical Problems without known medical cause[a]	11.Dependent 12.Lonely 25.NotGetAlong 27.Jealous 34.OutToGet 36.Accidents 38.Teased 48.NotLiked 62.Clumsy 64.PreferYoung 79.SpeechProb	9.MindOff 40.HearsThings 58.PicksSkin 60.SexPartsM 70.SeesThings 83.StoresUp 85.StrangeIdeas 100.SleepProb	18.HarmsSelf 46.Twitches 59.SexPartsP 66.RepeatsActs 76.SleepsLess 84.StrangeBehav 92.SleepWalk
Rule-Breaking Behavior		**Attention Problems**	**Aggressive Behavior**		**Other Problems**	
2.Alcohol 28.BreaksRules 43.LiesCheats 67.RunAway 73.SexProbs 82.StealsOther 96.ThinksSex 101.Truant 106.Vandalism	26.LacksGuilt 39.BadFriends 63.PreferOlder 72.SetsFires 81.StealsHome 90.Swears 99.Tobacco 105.UsesDrugs	1.ActsYoung 4.FailsToFinish 8.Concentrate 10.SitStill 13.Confused 17.Daydreams 41.Impulsive 61.PoorSchool 78.Inattentive 80.Stares	3.Argues 19.DemAtten 21.DestroyOther 23.DisobeySchl 57.Attacks 86.Stubborn 88.Sulks 94.Teases 97.Threaten	16.Mean 20.DestroyOwn 22.DisobeyHome 37.Fights 68.Screams 87.MoodChang 89.Suspicious 95.Temper 104.Loud	6.BMOut 15.CruelAnimal 44.BiteNail 55.Overweight 74.ShowsOff 93.TalkMuch 107.WetsSelf 109.Whining 113.OtherProb[b]	7.Brags 24.NoEatWell 53.Overeat 56h.OtherPhys 77.SleepsMore 98.ThumbSuck 108.WetsBed 110.WishOppSex

[a] Item 56 stands for "Physical Problems without known medical cause: other" and may not be filled in by the responder.
[b] Similarly, item 113 that stands for "Any problems that were not listed above" may be missing, for the same reason.

Total score for each subscale separately, total score for all items, change from baseline for the total score for each subscale separately and change from baseline for total score for all items will be summarized for each time point using descriptive statistics for continuous variables.

Total score is calculated as the sum of item scores.

If more than 20 items have missing values, excluding items 56h and 113, which may be left blank by structure (see Table 12 for details), then no scores will be provided. For calculating subscale scores, if one or more items, excluding items 56h and 113, are missing in a subscale, the subscale score will not be calculated.

Data will be listed.

In addition, individual plots will be provided.

8.12. Epworth Sleepiness Scale

The Epworth Sleepiness Scale (ESS) is a self-administered questionnaire composed of 8 questions that provide a measure of a patient's general level of daytime sleepiness (Johns 1991).

The ESS is administrated at screening, baseline, weeks 3, 7, 9, 12, 15/ET and 16/EOS.

The ESS is composed of 8 items. The responders are asked to rate their chances of falling asleep while engaged at 8 different activities, on a 4-point scale (0-3). Therefore, the total score can range from 0 to 24.

If at least 1 item has a missing value, then the entire questionnaire is not valid.

Total score of the 8 items and total scores change from baseline will be summarized by time point, using descriptive statistics. In the case of both caregiver and self-administered versions of the ESS are presented at the same visit, caregiver version will be used for patients who are 6 to 12 years old and self-administered version will be used for patients 13 years of age and older.

Total score is calculated as the sum of item scores.

Data will be listed.

8.13. Extrapyramidal Symptom Rating Scale

The Extrapyramidal Symptom Rating Scale (ESRS) consists of 2 subscales, ESRS I and ESRS II.

ESRS I and II are administrated at screening, baseline, weeks 3, 7, 9, 12, 15/ET and 16/EOS.

8.13.1. ESRS I

The first subscale (ESRS I) is a 7 item subjective questionnaire to evaluate parkinsonism, akathisia, dystonia and dyskinesia. The ESRS I is scored on for point scale (0=absent, 1=Mild, 2=Moderate, Severe=3) for each item.

Total score of the 7 items and total scores change from baseline will be summarized by time point, using descriptive statistics.

Total score will be the sum of the 7 items. Missing items will be replaced by the average of the remaining responses within each item.

Data will be listed.

8.13.2. ESRS II

The second subscale (ESRS II) is a 17-item questionnaire to evaluate parkinsonism and akathisia.

ESRS II consists of the following parts:

1. Tremor: scored on a 7-point item scale for 8 parts of the body. For each part of the body, frequency and amplitude are selected, assigning each one of the 8 items a score from 0=none to 6=severe. Therefore, the total score for tremor can ranged from 0 to 48.
2. Bradykinesia: scored on a 7-point item scale (0-6).
3. Gait and Posture: scored on a 7-point item scale (0-6).
4. Postural stability: scored on a 7-point item scale (0-6).
5. Rigidity: scored on a 7-point item scale (0-6) for 4 parts of the body. Therefore, the total score for rigidity can ranged from 0 to 24.
6. Expressive automatic movements: scored on a 7-point item scale (0-6).
7. Akathisia: scored on a 7-point item scale (0-6).

The total score of the 17 items can ranged from 0 to 102.

Total score of the 17 items and total scores change from baseline will be summarized by time point, using descriptive statistics for continuous variables.

Total score will be the sum of the 17 items. Missing items will be replaced by the average of the remaining responses within each item.

Data will be listed.

9. TOLERABILITY VARIABLES AND ANALYSIS

If more than 15% of the patients withdraw from the study before the end of the treatment period, the number of days until study discontinuation will be analyzed using Kaplan-Meier methodology with the ITT analysis set.

10. PLANNED INTERIM ANALYSIS

10.1. General

When approximately 50 patients have completed the study (including follow-up), an iDMC will perform an unblinded non-binding IA for futility based on centrally read MD- CRS part II total score.

A blinded SDR meeting will be performed prior to the IA to determine eligibility to the mITT analysis set and to classify missing data to MAR/MNAR.

Calculation of conditional power will be performed by an independent unblinded statistician, and the IA will be presented and discussed in a closed session of the iDMC. Release of randomization codes to the unblinded statistician will follow SOP GBP-RD-703.

Once the unblinded statistician and iDMC members are exposed to unblinded data, they will not be involved in decision making regarding the planned analysis of the study, or classification of missing data to MAR/MNAR.

Unblinded analysis datasets, TLGs, or simulation outputs will not be disclosed to the study team until the study is completed, database is locked and randomization codes are released for final analysis. The study team will be informed of the IA decision; communication format is described in the iDMC charter.

In order to minimize potential bias, investigators will not be informed on the parameters of the promising zone method, or on decision to increase the sample size.

The Statistical Analysis Plan and iDMC charter will be finalized before the IA is performed. Any changes in the main analysis of the primary and key secondary endpoints (eg, model specifications) will not be allowed after the IA.

10.2. Interim Analysis Procedure

The IA will include the following steps:

1. A blinded statistical data review meeting will be convened to review patient-blinded data from the patients that completed the study (including early termination). For patients that terminated early from the study, classification to MAR/MNAR will be performed. In addition, exclusion of patients from the mITT analysis set will be reviewed.
   The meeting will include blinded members only; iDMC members and the unblinded statistician will not take part in the discussions.
2. The sponsor will provide to the unblinded statistician the following data: disposition, demography, MD-CRS part II total scores (centrally read), the list of patients included in the mITT analysis set, and classification of missing data to MAR/MNAR. The unblinded statistician will have received the randomization codes from the randomization codes generator according to SOP GBP-RD-703.
3. The unblinded statistician will estimate the conditional power (see Section 10.4 for details) and provide the results to the iDMC.

4. The iDMC will apply futility rule (see Section 10.6) to provide recommendation to the sponsor to continue the study or stop the study for futility. This recommendation will be communicated with the sponsor while keeping the sponsor blinded, as specified in the iDMC Charter.
5. The sponsor will follow the iDMC recommendation for futility recommendations.

10.3. Data Cut for Interim Analysis

When the target number of patients (approximately 50 patients) complete the study (including follow-up; including patients that withdraw from the study) and the central reading of MD-CRS part II is available, the data will be cleaned for the IA. The IA data cut will not include data from patients that are ongoing.

Assessment of eligibility to the mITT analysis set and classification of missing data to MAR/MNAR will be performed in a blinded manner in a SDR meeting prior to the IA.

The following data will be used for the IA: disposition, demography, MD-CRS part II total scores (centrally read), the list of patients included in the mITT analysis set, and classification of missing data to MAR/MNAR.

10.4. Estimation of Conditional Power

The conditional power is the probability of rejecting the null hypothesis at the final analysis, assuming the alternative hypothesis is true given the data accumulated in the interim analysis. Conditional power will be estimated given the observed data, under the following assumptions:

- effect size (true difference and standard error) is as estimated in the observed data (point estimates)
- proportion of missing data MNAR in the TEV-50717 treatment group is as in the observed data
- proportion of patients excluded from the mITT analysis set is as in the observed data

Note: the conditional power is estimated based on the observed data and does not take into account the initial assumptions that were used for estimating the initial sample size.

The primary analysis is a 1-sided test of the null hypothesis H_0: $\delta \geq 0$ versus the alternative hypothesis H_1: $\delta < 0$, where δ is the difference between TEV-50717 and placebo in change from baseline in centrally-read MD-CRS part II total score (higher score represents a more severe movement disorder, thus negative change from baseline is a good outcome).

Let n_1 be the number of patients included in the mITT analysis set in the IA, and let n_2 be the incremental number of patients between the IA and the final analysis that will be included in the mITT analysis set, such that the total number of patients in the analysis will be $n = n_1 + n_2$.

Let z_1 be the Wald statistic at the IA. The conditional power is given by (Mehta and Pocock 2011):

$$\mathrm{CPow}(n_2|n_1, z_1) = 1 - \Phi\left(\frac{z_\alpha\sqrt{n} - z_1\sqrt{n_1}}{\sqrt{n_2}} - \frac{z_1\sqrt{n_2}}{\sqrt{n_1}}\right)$$

Calculation of the Wald statistic:

$$z_1 = \frac{-\hat{\delta}_1}{\mathrm{se}(\hat{\delta}_1)}$$

Where, $\hat{\delta}_1$ is the estimated mean difference between TEV-50717 and placebo and $\mathrm{se}(\hat{\delta}_1)$ is the estimated SE of the mean difference, both obtained from the primary analysis model (MMRM applying MAR/MNAR multiple imputation; see Section 6.2.3) fitted to the IA mITT data.

10.5. Sample Size

Let n_1^* be the total sample size at the IA (ITT analysis set). Assuming that the proportion of patients that will be excluded from the mITT analysis set for the primary analysis (ie, do not have at least 1 post-baseline central reading of MD-CRS part II total score) will be as estimated in the IA, the total sample size of the study is given by

$$n^* = \frac{n_1^*}{n_1} n = n_1^* + \frac{n_1^*}{n_1} n_2$$

The sample size recommendation will be presented in terms of the total sample size n^*.

10.6. Decision Rules

The following zone and corresponding actions are defined for all possible results of conditional power (CPow) is provided in Table 13.

Table 13: Decision Rules According to the Promising Zone Method for IA

Zone	Conditional Power	Decision
Futility	CPow <3%	Stop the study for futility

The futility decision is considered non-binding.

11. STATISTICAL SOFTWARE

All data listings, summaries, and statistical analyses will be generated using SAS® version 9.4 or later.

12. CHANGES TO ANALYSES SPECIFIED IN THE STUDY PROTOCOL

No changes to the analyses specified in the protocol are planned at the time of writing this SAP amendment.

13. REFERENCES

Achenbach TM. Advancing assessment of children and adolescents: commentary on evidence-based assessment of child and adolescent disorders. J Clin Child Adolesc Psychol 2005;34:541-7.

Achenbach TM, and Ruffle TM. The Child Behavior Checklist and Related Forms for

Assessing Behavioral/Emotional Problems and Competencies. Pediatrics in Review. 2000; 21(1): 265-271.

Battini R, Sgandurra G, Petacchi E, Guzzetta A, Di Pietro R, Giannini MT, et al. Movement Disorder-Childhood Rating Scale: reliability and validity. Pediatr Neurol 2008;39(4):259-65.

Johns MW. A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. Sleep 1991;14:50-55.

Little RJA, Rubin DB. Statistical Analysis with Missing Data, Second Edition. 2002, New York: John Wiley & Sons.

Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are promising: a practical guide with examples. Stat Med 2011;30(28):3267-84.

Rubin DB. Multiple Imputation for Nonresponse in Surveys. 1987, New York: John Wiley & Sons.

Heitjan F and Little RJA. Multiple Imputation for the Fatal Accident Reporting System. Applied Statistics 1991;40:13–29.

PEDI-CAT (Version 1.3.6) Development, Standardization and Administration Manual, Trustees of Boston University, under license to CREcare, LLC, 2011.

Hauser RA, Gordon MF, Mizuno Y, Poewe W, Barone P, Schapira AH, et al Minimal clinically important difference in Parkinson's disease as assessed in pivotal trials of pramipexole extended release. Parkinson's disease 2014, 467131. https://doi.org/10.1155/2014/467131

Hauser RA, Barkay H, Wilhelm A, Wieman M, Savola J-M, Gordon MF. Minimal clinically important change in abnormal involuntary movement scale score in tardive dyskinesia as assessed in pivotal trials of deutetrabenazine. Parkinsonism Relat Disord 2022;97:47-51.

Schenker N and Taylor JMG. Partially Parametric Techniques for Multiple Imputation. Computational Statistics and Data Analysis 1996;22:425-446.

Varni JW, Burwinkle TM, Berrin SJ, Sherman SA, Artavia K, Malcarne VL, et al. The PedsQL™ in pediatric cerebral palsy: Reliability, validity, and sensitivity of the Generic Core Scales and Cerebral Palsy Module. Developmental Medicine and Child Neurology 2006;48:442-49.

APPENDIX 1. MAR/MNAR JUMP TO REFERENCE MULTIPLE IMPUTATION METHODOLOGY

The following 3 steps will be performed to the do the multiple imputation analysis:

1. Create complete data sets using imputation for each subgroup.
2. Analyze the imputed complete data sets for the combined subgroups.
3. Combine analysis results into a final result.

Step 1

In order to induce monotonicity, an imputation model will first be used to impute missing values for MD-CRS part II total score week 9, 12, and 15 change values and will use age group at baseline and baseline MD-CRS part II total score to determine the scores (500 imputations will be created).

The SAS code for the imputation is as follows:

```
proc mi data=final out=mitv1 seed=987 nimpute=500;
 where arm='TEV-50717' & (missdm in ("MAR") or _9^=.);   /* 1st subgroup */
 class agegrp region;
 var agegrp region base _9;
 monotone regpmm; run;

proc mi data=final out=miplb1 seed=876 nimpute=500;
 where arm='PLACEBO' or (arm='TEV-50717' & missdm='MNAR' & _9=.);   /* 2nd subgroup */
 class agegrp region;
 var agegrp region base _9;
 monotone regpmm; run;

data mi1; set mitv1 miplb1; run;

proc sort data=mi1; by _imputation_; run;

proc mi data=mi1 out=mitv2 seed=9876 nimpute=1; by _imputation_;
 where arm='TEV-50717' & (missdm in (" 'MAR') or _12^=.);
 class agegrp region;
 var agegrp region base _9 _12;
 monotone regpmm; run;

proc mi data=mi1 out=miplb2 seed=8765 nimpute=1; by _imputation_;
 where arm='PLACEBO' or (arm='TEV-50717' & missdm='NMAR' & _12=.);
 class agegrp region;
 var agegrp region base _9 _12;
 monotone regpmm; run;

data mi2; set mitv2 miplb2; run;

proc sort data=mi2; by _imputation_; run;

proc mi data=mi2 out=mitv3 seed=98765 nimpute=1; by _imputation_;
 where arm='TEV-50717' & (missdm in (" 'MAR') or _15^=.);
 class agegrp region;
```

```
 var agegrp region base _9 _12 _15;
 monotone regpmm; run;


proc mi data=mi2 out=miplb3 seed=87654 nimpute=1; by imputation;
 where arm='PLACEBO' or (arm='TEV-50717' & missdm='NMAR' & _15=.);
 class agegrp region;
 var agegrp region base _9 _12 _15;
 monotone regpmm; run;
```

Step 2

The resultant complete data sets will be combined for the 2 subgroups and then analyzed using mixed-model, repeated-measures with the change in MD-CRS part II total score as the dependent variable (the primary analysis). The SAS code is as follows:

```
data all; set mitv3 miplb3;
     avisitn=9; chg=_9; output; avisitn=12; chg=_12; output; avisitn=15; chg=_15; output; run;

proc sort data=all; by _imputation_ descending arm avisitn; run;

proc mixed data=all order=data; by _imputation_;
   class usubjid arm avisitn agegrp region;
   model chg=arm avisitn arm*avisitn agegrp region base / ddfm=kr;
   repeated avisitn / type=un subject=usubjid;
   lsmeans arm*avisitn / diff cl;
   ods output diffs=diff(where=(avisitn=_avisitn) lsmeans=lsm; run;
```

Step 3

The results from the imputed data analysis will be combined using PROC MIANALYZE which will provide the LS mean and standard error of the LS mean for each treatment group at each visit as well as the LS mean difference, 95% CI and resultant p-value between the TEV-50717 and placebo groups at each visit. The SAS code is as follows:

```
proc sort data=diff; by avisitn arm; run;
proc sort data=lsm; by avisitn arm; run;

proc mianalyze data=diff; by avisitn arm;
 modeleffects estimate;
 stderr stderr;
 ods output ParameterEstimates=estdiff; run;

proc mianalyze data=lsm; by avisitn arm;
 modeleffects estimate;
 stderr stderr;
 ods output ParameterEstimates=estlsm; run;
```

APPENDIX 2. TIPPING POINT ANALYSIS

Let delta (shift) be the penalty to TEV-50717 group, ie. the worsening of effect for the patients in TEV-50717 group who early dropped out, compared to patients with full data, assuming MNAR. Delta will be in the range of 0 to 2 (or estimated treatment effect from primary analysis if it is larger than 2) in step of 0.2. The following 3 steps will be performed to do the multiple imputation analysis:

1. Create complete data sets using imputation for each shift in treatment effect.
2. Analyze the imputed complete data sets for each shift in treatment effect.
3. Combine analysis results into a final result for each shift in treatment effect.

Step 1:

The fully conditional specification (FCS) method will be used to impute missing values for MD-CRS part II total score change values and will use age group at baseline, treatment arm, and baseline MD-CRS part II total score to determine the scores (500 imputations will be created).

The SAS code for the imputation is as follows:

```
%macro midata(data=, smin=, smax=, sinc=, out=);

data &out; set _null_; run;

%let ncase=%sysevalf((&smax-&smin)/&sinc, ceil);

%do jc=0 %to &ncase;
 %let sj=%sysevalf(&smin + &jc*&sinc);

proc mi data=&data out=outmi seed=724375 nimpute=500;
 class arm agegrp region;
 fcs regpmm;
 mnar adjust (_9 /shift=&sj adjustobs= ( arm='TEV-50717'));
 mnar adjust (_12 /shift=&sj adjustobs=( arm='TEV-50717'));
 mnar adjust (_15 /shift=&sj adjustobs=( arm='TEV-50717'));

var arm agegrp region base _9 _12 _15; run;

data outmi; set outmi; shift=&sj; run;

data &out; set &out outmi; run;

%end;
%mend midata;

%midata(data=final, smin=, smax=, sinc=, out=out1);
```

Values should be added for smin, smax, and sinc. Per current definitions, these should be smin=0, smax=2.0 (or estimated treatment effect from primary analysis if it is larger than 2) and sinc=0.2.

Step 2

The resultant complete data sets will be analyzed using the same model as the primary analysis – see Appendix 1 - Step 2. The SAS code is as follows:

```
data all; set out1;
  avisitn=9; chg=_9; output; avisitn=8; chg=_8; output; avisitn=12; chg=_12; output; run;

proc sort data=all; by shift _imputation_ arm avisitn; run;

proc mixed data=all order=data; by shift _imputation_;
  class usubjid arm avisitn agegrp region;
model chg=arm avisitn arm*avisitn agegrp region base / ddfm=kr;
  repeated avisitn / type=un subject=usubjid;
  lsmeans arm*avisitn / diff cl;
  ods output diffs=diff(where=(avisitn=_avisitn )) lsmeans=lsm; run;
```

Step 3

The results from the imputed data analysis will be combined using PROC MIANALYZE which will provide the LS mean and standard error of the LS mean for each treatment group at each visit as well as the LS mean difference, 95% CI and resultant p-value between the TEV-50717 and placebo groups at each visit. The SAS code is as follows:

```
proc sort data=diff; by shift avisitn arm; run;
proc sort data=lsm; by shift avisitn arm; run;

proc mianalyze data=diff; by shift avisitn arm;
 modeleffects estimate;
 stderr stderr;
 ods output ParameterEstimates=estdiff; run;

proc mianalyze data=lsm; by shift avisitn arm;
 modeleffects estimate;
 stderr stderr;

 ods output ParameterEstimates=estlsm; run;
```