Statistical Analysis Plan

NCT Number: NCT04085172

Title: A Phase 4, Multicenter, 2-part Study Composed of a Randomized, Double-blind, Parallel-group, Placebo-controlled, Active-comparator, Dose-optimization Evaluation Followed by a 1-Year Open-label Evaluation to Assess the Safety and Efficacy of Guanfacine Hydrochloride Prolonged-release (SPD503) in Children and Adolescents Aged 6 to 17 Years With Attention-deficit/Hyperactivity Disorder

Study Number: SPD503-401

Document Version and Date: Version 3.0, 21 Jul 2025

Certain information within this document has been redacted (ie, specific content is masked irreversibly from view) to protect either personally identifiable information or company confidential information.

STATISTICAL ANALYSIS PLAN

Study Number: SPD503-401

Study Title: A Phase 4, Multicenter, 2-part Study Composed of a Randomized, Double-blind, Parallel-group, Placebo-controlled, Active-comparator, Dose-optimization Evaluation followed by a 1-Year Open-label Evaluation to Assess the Safety and Efficacy of Guanfacine Hydrochloride Prolonged-release (SPD503) in Children and Adolescents aged 6 to 17 Years with Attention-deficit/Hyperactivity Disorder

Phase: IV

Version: 3.0

Date: 21 Jul 2025

Status: Final

Prepared by:

[redacted], Statistical and Quantitative Sciences, Takeda

Based on:

Protocol Version: Amendment 8.0

Protocol Date: 12-Apr-2024

REVISION HISTORY

Version	Approval Date	Primary Rationale for Revision
1.0	09Aug2021	N/A
2.0	21Jun2024	Align with substantial changes in Protocol Amendment 7 and address issues identified during dry run
3.0	21Jul2025	Address clinical input identified during dry run #2

TABLE OF CONTENTS

ABBREVIATIONS

ADHD	attention-deficit/hyperactivity disorder
ADHD-RS-5	ADHD-Rating Scale-5
AE	adverse event
ALT	alanine aminotransferase
APRS	Academic Performance Rating Scale
AST	aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
BMI	body mass index
BPRS-C-21	Brief Psychiatric Rating Scale for Children (21-item questionnaire)
BUN	blood urea nitrogen
C3PS	Conners 3 Parent Short Form
CANTAB	Cambridge Neuropsychological Test Automated Battery
CDC	Center for Disease Control
CGI-I	Clinical Global Impression–Improvement
CGI-S	Clinical Global Impression–Severity
CHIP-CE:PRF	Child Health and Illness Profile – Child Edition: Parent Report Form
CMH	Cochran-Mantel-Haenszel test
COVID-19	Coronavirus Disease 2019
CPK	creatine phosphokinase
CRF	case report form
CSR	clinical study report
C-SSRS	Columbia-Suicide Severity Rating Scale
DMS	Delayed Matching to Sample (CANTAB task)
ECG	electrocardiogram
EF	executive function
EMA	European Medicines Agency
EU	European Union
FAS	full analysis set
FDA	Food and Drug Administration
GGT	γ-glutamyl transferase
ICH	International Conference on Harmonisation
IMP	investigational medicinal product
IRT	Interactive Response Technology
IVRS	Interactive Voice Response System
K-SADS-PL	Kiddie-Schedule for Affective Disorders-Present and Lifetime Version
LDH	lactate dehydrogenase
LLN	lower limit of normal
LS	least squares
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed-Effects Model for Repeated Measures
QD	once daily
PASS	postapproval safety study
PCI	potentially clinically important

PDSS	Pediatric Daytime Sleepiness Scale
PPS	per-protocol set
PRO	patient-reported outcome
QD	once daily
RTI	Reaction Time/5-Choice Reaction Time (CANTAB task)
RVP	Rapid Visual Information Processing (CANTAB task)
SAE	serious adverse event
SAP	statistical analysis plan
SDB	standard database
SI	International System of Units
SST	Stop Signal Task (CANTAB task)
SWM	Spatial Working Memory (CANTAB task)
TEAE	treatment-emergent adverse event
TLGs	tables, listings, and graphs
TSH	thyroid stimulating hormone
UKU	Udvalg for Kliniske Undersøgelser
ULN	upper limit of normal
WHODrug	World Health Organization Drug Dictionary

1.0 OBJECTIVES, ENDPOINTS AND ESTIMANDS

1.1 Objectives

1.1.1 Primary Objective

The primary safety objective of this study is to evaluate the comparative long-term safety of TAK-503 (formerly known as SPD503) treatment in children and adolescents aged 6 to 17 years diagnosed with ADHD for whom stimulants are not suitable, not tolerated, or shown to be ineffective:

- To evaluate TAK-503 compared with atomoxetine after 12 months of once daily (QD) treatment on psychomotor speed and attention as measured by the Cambridge automated neuropsychological test battery (CANTAB) reaction time (RTI) task, using the mixed-effects model for repeated measures (MMRM). The effect of TAK-503 on cognition will be assessed and interpreted on the totality of the data.

A primary efficacy objective is not applicable for this postapproval safety study (PASS).

1.1.2 Secondary Objective(s)

Secondary safety objectives

- Cognitive domain, sustained attention as measured by the CANTAB Rapid Visual Information Processing (RVP) task
- Cognitive domain, Spatial Working Memory (SWM), a component of executive function (EF), as measured by the CANTAB SWM task between errors
- Cognitive domain, response control/inhibition as measured by the CANTAB Stop Signal Task (SST)
- Cognition domain, recognition memory as measured by the CANTAB Delayed Matching to Sample (DMS) task
- Sexual maturation as measured by Tanner stage
- Growth as measured by weight, height, and body mass index (BMI)
- Incidence of treatment-emergent adverse events (TEAEs)
- Vital sign and electrocardiogram (ECG) results
- Psychiatric symptoms as measured by the Brief Psychiatric Rating Scale for Children (BPRS-C) total score and factors for Depression, Anxiety, Psychomotor Excitation, Behavior Problems, Withdrawal, Thinking Disturbance, and Organicity
- Suicidal ideation/behavior as measured by the Columbia-Suicide Severity Rating Scale (C-SSRS)
- Asthenia/Lassitude/Increased Fatigability, Sleepiness/Sedation, Increased Duration of Sleep, and Orthostatic Dizziness based on the ratings of specific Udvalg for Kliniske Undersøgelser (UKU) scale items
- Sedative effects as measured by subject ratings on the Pediatric Daytime Sleepiness Scale (PDSS)

Secondary efficacy objectives:

- Attention-deficit/hyperactivity disorder symptoms as measured by the investigator-administered ADHD-Rating Scale-5 (ADHD-RS-5) total score and hyperactivity/impulsivity and inattention subscale scores
- Global clinical measurement of ADHD improvement as measured by Clinical Global Impression–Improvement (CGI-I) using the Clinical Global Impression–Severity (CGI-S) to establish baseline
- Function as measured by the Child Health and Illness Profile – Child Edition: Parent Report Form (CHIP-CE:PRF) domains of satisfaction, comfort, resilience, risk avoidance, and achievement. The effect of TAK-503 on all subdomains will be summarized, including the subdomain satisfaction with self, using parental ratings of the child's self-esteem.
- Assessment of behavioral, social, and academic issues with the Conners 3 Parent Short Form (C3PS) Total Score and the Learning Problems and Executive Functioning subscale scores

1.2 Endpoints

1.2.1 Primary Endpoint(s)

- The primary safety endpoint will be the change from baseline in the CANTAB RTI task.

1.2.2 Safety Endpoints

- Secondary safety endpoints will include the following:
    - CANTAB tasks: RVP, SWM between errors, DMS, and SST
    - Tanner stage, weight, height, BMI
    - Vital signs (BP and pulse) and ECG results, and overall safety events
    - BPRS-C total score and scales for Depression, Anxiety, Psychomotor Excitation, Behavior Problems, Withdrawal, Thinking Disturbance, and Organicity
    - C-SSRS
    - Specified UKU side effect rating scale items: Asthenia/Lassitude/Increased Fatigability, Sleepiness/Sedation, Increased Duration of Sleep, and Orthostatic Dizziness
    - PDSS

1.2.3 Efficacy Endpoints

A primary efficacy endpoint is not applicable for this PASS.

Secondary efficacy endpoints will include the following:

- ADHD-RS-5 total score and subscale scores for hyperactivity/impulsivity and inattention domains
- CGI-I, calculated from CGI-S

- CHIP-CE:PRF
- C3PS total Score and scores for Learning Problems and Executive Functioning subscales (for subjects enrolled with or after Amendment 5)

1.2.4 Other Endpoints

- Academic Performance Rating Scale (APRS); for subjects enrolled before Amendment 5 who have a baseline APRS assessment in Part A

1.3 Estimand(s)

The estimand framework for the superiority analysis of the primary safety endpoint is illustrated below in Table 1.

Table 1 Estimand Framework

Estimand: [Primary]

	Attributes				
Definition	**Treatment**	**Population**	**Variable (or Endpoint)**	**Strategy for Addressing Intercurrent Event**	**Population-Level Summary**
The primary estimand is the treatment effect of TAK-503 or Atomoxetine (separately) compared to Placebo at Week 18A in targeted patient population	TAK-503 or Atomoxetine QD as single intervention	Children and adolescents aged 6 to 17 years diagnosed with ADHD for whom stimulants are not suitable, not tolerated, or shown to be ineffective and who complete 18 weeks of DB treatment for ADHD	Change from baseline to Week 18A in CANTAB RTI.	*Intercurrent Event:* Treatment Discontinuation prior to Week 18A *Strategy:* Subjects will be included in MMRM analysis as long as Week 10A CANTAB RTI is present *Intercurrent Event:* COVID-19 impact *Treatment-Policy Strategy:* Data from visits impacted by COVID-19 are sufficient for completion of the CANTAB RTI assessment and therefore will be analyzed regardless of such intercurrent events	Difference in variable means between TAK-503 or Atomoxetine and Placebo.

2.0 STUDY DESIGN

Study SPD503-401 is a phase 4, multicenter, dose-optimization PASS and will be conducted in 2 parts: Study Parts A and B. Study Part A will be a randomized, double-blinded, parallel-group, placebo- and active comparator-controlled, 3-treatment arm safety and efficacy evaluation of TAK-503. Pediatric subjects with ADHD will be randomized 1:1:1 among TAK-503, atomoxetine, and placebo treatment arms for the first 18 weeks of double-blinded treatment and evaluation. At the end of the first 18 weeks (Week 18A/Visit 11A), all subjects will roll over to Study Part B directly for an additional 52 weeks of open-label TAK-503 treatment. Once 120

subjects have entered into Study Part B, only subjects in the placebo arm in Study Part A will enter Study Part B, to assure that all placebo arm subjects have the opportunity to enter into the 1-year TAK-503 open-label treatment arm.

An illustration of the study design is available below in Figure 1. Study Part A consists of the following 5 periods elapsing up to 23 weeks.

1. Screening/washout: Day -35A to Day -3A (up to 5 weeks).

2. Dose-optimization period: Investigational medicinal product (IMP) will be dose-optimized from Week 0A through Week 4A for children aged 6 to 12 years and through Week 7A for adolescents aged 13 to 17 years.

3. Dose-maintenance period: Subjects will continue with QD dosing from Week 5A for children and Week 8A for adolescents through Week 18A.

4. Dose-taper period: Scheduled 3-week fixed-dose taper of IMP; All subjects randomized in the TAK-503 or atomoxetine treatment arms will undergo a scheduled 3-week dose taper to ensure proper downward dose titration of TAK-503 at the end of the dose-maintenance period or early termination (ET).

5. Follow-up period: Follow-up visit 1 week after the last IMP tapered dose

    - Subjects randomized to TAK-503 or atomoxetine treatment arms will participate in a follow-up visit and begin rescreening/washout procedures for Study Part B.

    - Subjects randomized to the placebo treatment arm can begin baseline visit assessments for Study Part B and skip the follow-up visit after completion of Study Part A assessments at Visit 11A (Week 18A).

Before enrollment in Study Part A, subjects will be screened based on inclusion/exclusion criteria to establish eligibility for study participation. Approximately 25% of all subjects will be aged 13 to 17 years. Approximately 25% of all subjects will be female. Subjects who meet eligibility requirements will undergo a medication washout period, if applicable. In Study Part B, subjects randomized to active IMP in Study Part A must undergo a washout period of ≥30 days between the last IMP tapered dose in Study Part A and the first TAK-503 dose in Study Part B. In Study Part A, subjects with a baseline ADHD-RS-5 total score ≥28 and CGI-S score of ≥4 will be eligible for enrollment. During the dose-optimization period, clinic visits will be scheduled every 7 days to assess IMP safety and tolerability and to allow investigators to titrate the IMP to an optimal dose that maximizes potential benefits while minimizing risk of TEAEs. If necessary, the investigator will be allowed to lower the subject's dose once during the dose optimization period in Study Parts A and B.

Study Part B will consist of 5 periods:

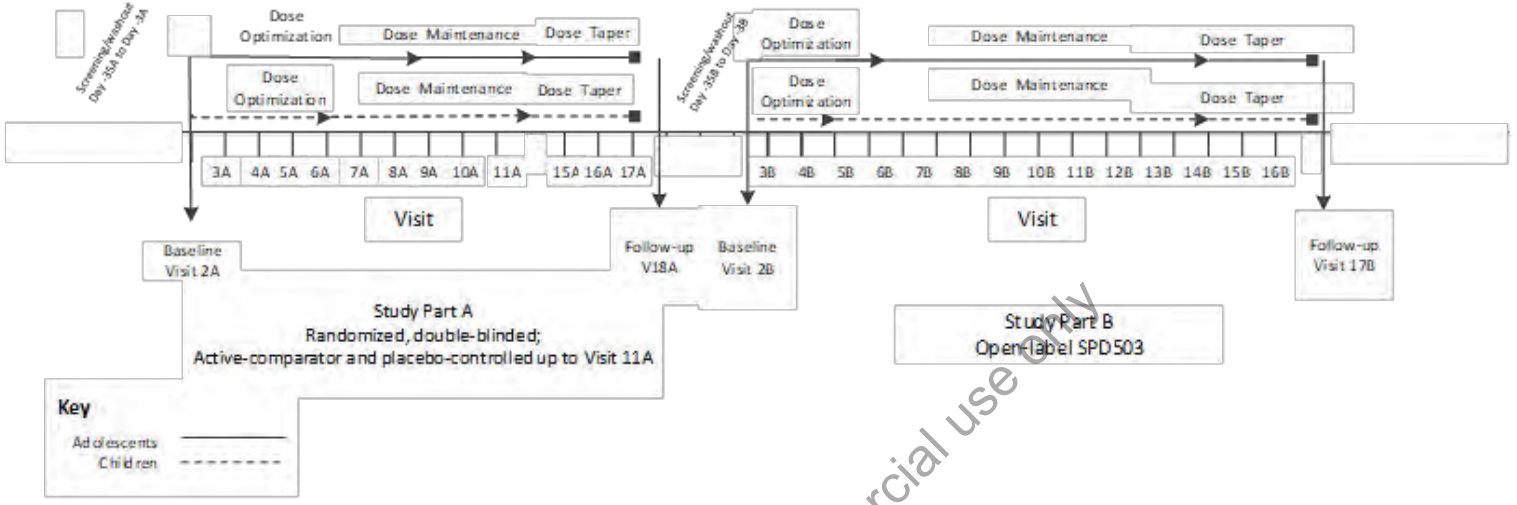
1. Screening/washout: Day -35B to Day-3B (up to 5 weeks)

2. Dose-optimization period: TAK-503 will be dose-optimized from Week 0B through Week 4B for children aged 6 to 12 years and through Week 7B for adolescents aged 13 to 17 years.
3. Dose-maintenance period: Dose-optimized TAK-503 will be administered QD from Week 5B for children and from Week 8B for adolescents through Week 49B for both age groups.
4. Dose-taper period: Scheduled 3-week downward-titration of TAK-503 dose through Weeks 50B, 51B, and 52B
5. Follow-up period: Last dose of TAK-503 to the follow-up visit at Week 53B

During the Study Part A follow-up visit, subjects will undergo a partial rescreening of study criteria to confirm eligibility for continued participation in Study Part B. Subjects who complete Study Part A should roll over into Study Part B without a gap except for washout. Subjects (except those randomized to the placebo group) must undergo a washout period of ⩾30 days between the last IMP tapered dose in Study Part A and the first dose of TAK-503 in Study Part B. Eligibility to continue in Study Part B will be confirmed at Visit 2B, Study Part B baseline. Once 120 subjects have entered into Study Part B, only subjects in the placebo arm in Study Part A will enter Study Part B, to assure that all placebo arm subjects have the opportunity to enter into the 1-year TAK-503 open-label treatment arm.

Each subject's last visit will be 7 (+2) days after the last TAK-503 dose in Study Part B, to measure blood pressure and heart rate, and to follow up on concomitant medications and safety assessments, including TEAEs, which may have been ongoing at the previous visit. Treatment-emergent adverse events reported up to the time of the follow-up visit will be documented and appropriate follow-up will continue until all safety concerns are resolved as judged by the investigator. A full accounting of the study schedules for Study Parts A and B can be found in Tables 1 and 2 of the Protocol, respectively.

Figure 1 Study Design Flow Chart



Note: SPD503 is now known as TAK-503.

3.0 STATISTICAL HYPOTHESES AND DECISION RULES

3.1 Statistical Hypotheses

For the superiority analysis of the primary safety endpoint comparing each active arm (separately) against placebo, the corresponding 1-sided null (H_0) and alternative (H_1) hypotheses are:

- H_0: $\mu_{(\text{TAK-503/Atomoxetine})} - \mu_{(\text{Placebo})} \geq 0$ (i.e. TAK-503/Atomoxetine is not superior to placebo)

versus

- H_1: $\mu_{(\text{TAK-503/Atomoxetine})} - \mu_{(\text{Placebo})} < 0$ (i.e. TAK-503/Atomoxetine is superior to placebo)

where $\mu_{(\text{treatment})}$ is the mean change from baseline to Week 18A in reaction time (msec) for the CANTAB RTI task for the corresponding treatment (a lower score indicates greater improvement).

3.2 Statistical Decision Rules

As suggested by the European Medicines Agency (EMA), using data up to 18 weeks, the superiority to placebo comparisons including TAK-503 and atomoxetine, separately, will be performed using mixed-effects model for repeated measures (MMRM) as specified for the primary endpoint (change from baseline in CANTAB RTI). If TAK-503 is statistically superior to placebo, then a negative impact on cognition can be excluded. If atomoxetine is statistically superior to placebo, then the assay sensitivity is demonstrated.

If neither TAK-503 or atomoxetine is superior to placebo, then the noninferiority between TAK-503 and placebo at 18 weeks will be assessed using the margin of mean change from baseline in placebo arm. The 1-sided 97.5% confidence interval (upper bound) for the difference between TAK-503 and placebo will be derived from the MMRM. If the upper bound of the 1-sided 97.5% CI for the difference lies entirely below the margin (in msec), then it will be concluded that TAK-503 is noninferior to placebo for the primary endpoint.

In addition, the effect of TAK-503 and atomoxetine will be compared at 1 year (Week 49A) in a descriptive fashion using MMRM, without the performance of inferential hypothesis testing. Prior to elimination of the 18 to 49 week portion of Study Part A in Protocol Amendment 7, data for CANTAB domains (and other assessments) were collected through 49 weeks in Study Part A for active arm subjects.

All analyses will be performed at the end of the study.

3.3 Multiplicity Adjustment

No multiplicity adjustments are planned for this study.

4.0 SAMPLE-SIZE DETERMINATION

In this safety study, a descriptive analysis of TAK-503 and atomoxetine will be performed after 12 months of double-blinded treatment using results from the CANTAB RTI task. The effect of TAK-503 on cognition will be assessed and interpreted on the totality of the data.

The sample size was estimated using the assumed difference of 0 msec with a noninferiority margin of 30 msec and a common standard deviation of 63.4 for the comparison between TAK-503 and atomoxetine at Week 49A on the RTI in Study Part A. The assumed common standard deviation on RTI is based on a prior study SPD503-206: a randomized, double-blinded, placebo-controlled efficacy study of TAK-503 in children and adolescents with ADHD. The assumed margin of 30 msec is close to the value (29.3 msec) of change from baseline in reported results from an adolescent study with atomoxetine (Shang and Gau, 2012). The assumed margin was used for sample size only and will not be applied in the noninferiority evaluation.

A total of 216 subjects, 72 subjects per arm, will be required to provide 80% power for a 2-sided 95% CI. Expecting a nonevaluable rate of 25%, including postrandomization dropout, during the double-blinded period, the randomization target has been set to 288 subjects total or 96 subjects per treatment arm.

As of Protocol Amendment 7, TAK-503 and atomoxetine will be evaluated after up to 12 months of double-blinded treatment using results from the CANTAB RTI task via the MMRM; however, inferential hypothesis testing via the noninferiority framework at the Week 49A timepoint will no longer be performed due to challenges with retaining an evaluable sample size that would allow for sufficient power for this comparison; superiority to placebo comparisons of TAK-503 and atomoxetine at the 18-week timepoint, separately, will still be performed, as in the original protocol.

The number of subjects projected to participate in Study Part B is 120, which will be sufficient to have at least 90% power to obtain a 95% confidence interval for the change from baseline RTI value that excludes the value 0, given the assumed response values for TAK-503 based on Study SPD503-206 (mean=20.7, SD=63.11).

5.0 ANALYSIS SETS

5.1 Randomized Set

The randomized set is defined as all randomized subjects in Study Part A. Any subject who is not randomized will be considered a screen failure.

5.2 Double-Blind Safety Set

The double-blind safety set will be defined as all randomized subjects in Study Part A who receive ≥1 IMP dose. The double-blind safety set will be used for safety analysis (except for CANTAB domains) for Study Part A.

5.3 Open-Label Safety Set

The open-label safety set will be defined as all subjects who receive ≥1 dose of TAK-503 in Study Part B. All safety and efficacy presentations for Study Part B will be based on the open-label safety set. Tables using the open-label safety set will include grouping by treatment received during the double-blind study part.

5.4 Full Analysis Set

The full analysis set (FAS) will be defined as all subjects in the double-blind safety set with ≥1 postbaseline CANTAB assessments in Study Part A. This analysis set will be used for the analysis of all efficacy endpoints, as well as analyses of the primary safety endpoint and CANTAB domains, for Study Part A.

5.5 Per-Protocol Set

The per-protocol set (PPS) will be defined as all subjects in the FAS who complete Study Part A and were deemed protocol-compliant; For subjects enrolled following implementation of Protocol Amendment 7 (given the elimination of the 18 to 49 week portion of Study Part A to reduce patient burden), Study Part A completion for all treatment arms is defined as completion of the Week 18A visit.

To be protocol-compliant, a subject will not have had any significant protocol deviations during the study that could affect the assessment of the primary safety endpoint, the CANTAB RTI task. Specific protocol deviations used to define this analysis set will be finalized prior to the database lock. This analysis set will be used for the sensitivity analyses of the primary safety endpoint and all CANTAB domains for Study Part A.

6.0 STATISTICAL ANALYSIS

6.1 General Considerations

The “baseline” assessment for Study Part A will be defined as the last observed value before the first dose of blinded IMP. For subjects who enter Study Part B, the “baseline” assessment for analyses using the open-label safety set is the last observed value before the first dose of open-label TAK-503, including consideration of values from Study Part A, unless otherwise specified.

All confidence intervals, statistical tests, and resulting p-values will be reported as 2-sided and will be assessed at the nominal α=0.05 significance level without adjustment for multiplicity, unless otherwise stated.

Where appropriate, variables will be summarized descriptively by study visit. Unless otherwise specified for the particular assessment, a windowing convention will be used to determine the analysis value for a given study visit for observed data analyses, as defined in Section 9.2.3.

For the categorical variables, the count and proportions of each possible value will be tabulated by treatment arm. The denominator for the proportion will be based on the number of subjects who provided non-missing responses to the categorical variable.

For continuous variables, the number of subjects with non-missing values, mean, median, SD, minimum, and maximum values will be tabulated. Change from baseline will be calculated by subtracting the baseline value from the post baseline value. Percent change from baseline will be calculated as:

((Observed measurement at Post-Baseline Visit – Observed measurement at Baseline Visit) / Observed Measurement at Baseline Visit) * 100%

This is a multi-center study. However, due to the large number of centers with low numbers of subjects, center will not be included as a factor in statistical models. Summary tables for select endpoints may be produced by center.

Assessment scores will be calculated if the number of items with missing or invalid data is less than or equal to 20% of the number of total items (unless otherwise specified for the particular assessment). Otherwise, the assessment score will be set to missing. Unless otherwise stated, incomplete data that are resulted from either early study termination or unavailability will be left as missing.

For both Study Parts A and B, outputs for all analysis sets will be grouped by "Placebo", "TAK-503", "Atomoxetine", and "Total" arms, unless otherwise specified. Figures corresponding to the single-arm, open-label Study Part B will display only the "Total" arm, representing all subjects in the open-label safety set.

A masked data review of tables, listings, and figures presented with surrogate treatment codes will be conducted prior to unblinding of subjects' treatment assignments following study completion. This review will assess the accuracy and completeness of the study database, subject evaluability (including definition of analysis sets), and appropriateness of the planned statistical methods.

6.1.1 Handling of Treatment Misallocations

Subjects in the double-blind safety set, open-label safety set and PPS will be analyzed according to treatment received during the double-blind study part. Subjects in the FAS will be analyzed in the treatment arm to which they were randomized during the double-blind study part, regardless of treatment received.

6.1.2 Analysis Approach for Continuous Variables

For relevant continuous variables, the treatment effect will be analyzed via a mixed-effects model for repeated measures (MMRM), which will include fixed effects for treatment arm, visit, sex (male or female), age group (6 to 12 years or 13 to 17 years), region (Europe or US), and the interaction of treatment with visit, with visit as the repeating factor and subject as a random effect, the corresponding baseline value as a covariate, and the interaction between baseline

value and visit adjusted in the model. Only observations from analysis visits corresponding to scheduled visits (including scheduled Part A visits for which data were collected prior to those visits being eliminated under Protocol Amendment 7) of a given assessment will be included in the model.

An unstructured covariance matrix will be used to model the correlation among repeated measurements; parameter estimates will be calculated using restricted maximum-likelihood estimation (REML). If the model fails to converge, the following structures will be considered, in the specified order, until convergence is achieved: Toeplitz with heterogeneity, autoregressive with heterogeneity by visit, compound symmetry with heterogeneous variances by visit, autoregressive, Toeplitz, and compound symmetry without heterogeneous variances by visit. Degrees of freedom will be estimated using the Kenward-Roger approximation.

The least squares (LS) mean estimates of the change from baseline for each treatment arm, the differences between the LS means across treatment arms, and 95% confidence intervals and p-values associated with the differences will be reported at each scheduled visit of Study Part A (including scheduled Part A visits for which data were collected prior to those visits being eliminated under Protocol Amendment 7), as applicable.

Since the elimination of the 18 to 49 week portion of Study Part A under Protocol Amendment 7 was implemented at a distinct timepoint, impacting active arm subjects equally and simultaneously (eg, systemic bias), subsequent "missing" assessment data at these eliminated timepoints will be assumed to be missing at random, and analyses will be performed using MMRM to handle potential missing data. No imputation will be performed in this regard.

Sample coding for generating the MMRM in SAS is below. This sample SAS code is suggestive, and final SAS code used in production work needs to be appropriately applied to the dependent and independent variables being analyzed and to be validated by an independent statistician(s). The native SAS outputs for the execution of these analyses (i.e. PROC MIXED) will be provided as documentation in section 16.1.9.2 of the clinical study report.

```
PROC MIXED data=cantab;
        CLASS trt01pn visit region sex agegrp subjid;
        MODEL CFB= baseline trt01pn visit region sex agegrp trt01pn*visit
        baseline*visit/ DDFM=KR;
        REPEATED visit/ subject=subjid type=UN;
RUN;
```

To estimate the difference between Atomoxetine and Placebo at Week 18A, as an example, the below code may be used:

```
ESTIMATE 'A-P Difference at Week 18'
```

```
TRT01PN                 -1 0 1

TRT01PN*VISIT           0 -1

                        0 0

                        0 1   / cl;
```

The continuous variables to which the above analysis will be applicable, for Study Part A only, are listed below:

- ADHD-RS-5: subscale and total scores
- CHIP-CE:PRF: global, domain and subdomain raw scores
- CANTAB RTI: Reaction time, movement time, for both simple and five-choice variants
- CANTAB RVP: A', mean response latency, probability of hit
- CANTAB SWM: Strategy, Total Errors
- CANTAB SST: Stop Signal Reaction Time, Direction Error (Go trials), Direction Error (Stop trials), Median Reaction Time (All Go trials), Missed Trials
- CANTAB DMS: Mean Choices to Correct, Mean Correct Latency, Proportion of Correct Responses
- BPRS-C total score
- PDSS total score
- C3PS total score and subscale t-scores (Learning Problems, Executive Functioning)

For comparisons of CANTAB domains through Week 18A which use the PPS (as part of the sensitivity analysis), only PPS subjects who complete Week 18A (for all treatment arms) will be included in the MMRM. For the Week 49A comparison using the PPS, only PPS subjects who complete Week 49A will be included in the MMRM (ie will exclude placebo subjects and subjects enrolled under Protocol Amendment 7 or later). Analyses using the FAS will only filter for subjects who have ≥ 1 post-baseline CANTAB assessment at Week 10A or later.

For all other continuous endpoints listed above, all subjects in the relevant analysis set will be included in the respective analysis model.

6.1.3 Analysis Approach for Binary Variables

All binary efficacy variables in this trial will use the following analysis method, unless stated otherwise in the section specific to an endpoint: Cochran Mantel Haenzsel tests, stratified by age group (6 to 12, 13 to 17) and sex, will be performed at each post-baseline visit thru Week 18A to determine differences between the proportions in the placebo vs TAK-503 and the placebo vs. atomoxetine treatment arms. Proportions, along with 95% CI based off the Clopper Pearson method, will be reported for each treatment arm, along with differences in proportions for the treatment comparisons. The p-values, based off the CMH row-mean score test (van Elteren test), will be provided. Only observations from analysis visits corresponding to scheduled visits (including scheduled Part A visits for which data were collected prior to those visits being eliminated under Protocol Amendment 7) of a given assessment will be included in the model.

Similarly, these tests will be performed at all post-baseline visits thru Week 49A for the comparison between TAK-503 and atomoxetine, considering subjects with available data at a given visit. The native SAS outputs for the execution of these analyses (ie PROC FREQ) will be provided as documentation in section 16.1.9.2 of the clinical study report.

The binary variables to which the above analysis will be applicable, for Study Part A only, are listed below:

- CGI-I Improvement (Percentage with CGI-I of 1 or 2)
- Responder (Percentage with response, defined as reduction from baseline ADHD-RS-5 total score of ≥30% and CGI-I of 1 or 2 at a given post-baseline visit)

6.1.4 Handling of Remote Visits and Assessments

If a subject is unable to participate in a visit on site due to unavoidable dire circumstances (eg, a widespread disease outbreak such as the COVID-19 pandemic or natural disaster), remote visits may be conducted by phone (eg, collection of AEs and monitoring), video conferencing (telehealth/telemedicine), or qualified site staff or qualified designees visiting the participant's residence. During the dose maintenance phase, subjects may participate in up to 2 consecutive remote visits. Note that this is not a cumulative limit on the total number of remote visits. A subject may have more than 2 remote visits throughout the entire course of the study, but cannot have 3 remote visits in a row (unless required due to unavoidable dire circumstances, eg, a widespread disease outbreak such as the COVID-19 pandemic or natural disaster). Details of these impacted visits will be recorded in the eCRF.

If data permit, to assess the consistency of results collected during such assessments, sensitivity analyses will be conducted to determine the impact of alternative methods of data collection, as described in Section 6.6.1.3.

6.2 Disposition of Subjects

Disposition of Study Part A subjects will include summaries for subjects screened and randomized. Subjects screened will be summarized by tabulating, i.e. counts and percentages, total screens, randomized, screen failures, and if data permit, subjects that rescreen. For subjects that screen fail, the reasons for failure (including COVID-19) will also be tabulated. Reasons for rescreen will be tabulated, if data permit.

Subjects in the randomized set will be summarized by tabulating the variables listed below, using randomized DB treatment arms as well as all subjects overall:

- subjects by site, country and region
- subjects randomized but not treated
- subjects who completed Week 18A visit
- subjects who completed Week 49A visit (for active treatment arms only)
- subjects who completed Week 52A visit (for active treatment arms only)
- subjects who completed treatment of Study Part A
- subjects who discontinued from treatment during Study Part A (by reason)
- subjects who completed Study Part A
- subjects who discontinued from study during Study Part A (by reason)
- subjects impacted by COVID-19 (by visit impact), as identified by the COVID-19 Impact CRF
- subjects who continued to Part B of study (including breakdown of enrolled but not treated with OL dose, and treated with OL dose)

As Protocol Amendment 7 led to the elimination of the 18 to 49 week portion of Study Part A, subjects enrolled in the study at the time of this protocol implementation who complete treatment/study through Week 18A will still be considered treatment/study completers for Study Part A, even if they undergo an early termination visit after Week 18A. The timing of the completion (e.g. before or after Protocol Amendment 7) will be presented, as appropriate.

For Study Part B, a summary of screen failures will be presented. In addition, subjects in the open-label safety set will be summarized by tabulating the variables listed below, grouped by randomized DB treatment arms as well as all subjects overall:

- subjects by site, country and region
- subjects who completed Study Part B
- subjects who discontinued from study during Study Part B (by reason)
- subjects impacted by COVID-19 (by visit impact), as identified by the COVID-19 Impact CRF

A by-subject listing of study completion information, including the timing and reason for study discontinuation, will be presented for all subjects. This listing will also indicate if a subject's Study Part A participation was impacted by implementation of Protocol Amendment 7.

In addition, the number and percentage of subjects in the FAS experiencing a protocol deviation during Study Part A will be summarized by deviation severity and category. If appropriate, an additional summary of COVID-19-related protocol deviations may be presented. An additional table will be presented filtered for significant (major or critical) protocol deviations occurring during Study Part A. These tables will be repeated for Study Part B, using the open-label safety set.

A by-subject listing will be presented for all protocol deviations, using the FAS, with a flag to indicate if the deviation resulted in exclusion from the PPS.

A by-subject listing of all participants impacted by the COVID-19 pandemic and a description of how the individual's participation was altered, as recorded by the COVID-19 Impact CRF, will be presented.

6.3 Demographic and Other Baseline Characteristics

6.3.1 Demographics and Baseline Characteristics

Descriptive statistics (mean, median, SD, minimum, and maximum values) will be provided for continuous variables (age, height, weight, and body mass index [BMI], baseline CANTAB RTI (five-choice), baseline ADHD-RS-5 total score). The number and percentage of subjects will be provided for the following categorical variables: race, ethnicity, sex, age group (6 to 12, 13 to 17), derived age group (2 to 11, 12 to 17), country, region (Europe, US), history of prior stimulant medication (yes, no), Tanner Stage Scores for Males/Females (I to V), CGI-S score at baseline.

For Study Part A, age will be calculated as the difference between date of birth (DOB) and date of informed consent (DINFC), truncated to months, using the following formula:

- age = floor((intck('month',DOB,DINFC) -1) / 12)

For Study Part B, age will be calculated in an analogous manner using the Part B screening date.

Age group (6 to 12, 13 to 17) may differ between Study Part A and Study Part B within a subject, and will be presented as collected in IRT using the "age group at randomization" and "age group for enrollment" variables, respectively. The relevant age group corresponding to each study part will be selected for subgroup analyses.

BMI will be calculated as follows:

- BMI = (weight in kg x 10,000)/(height in cm)^2

Summary statistics will be generated overall and by treatment arm based on the double-blind safety set. No inferential statistics will be presented. This table will be repeated for the FAS, if

there exists a greater than 5% discrepancy in subject count compared to the double-blind safety set. An additional set of tables, stratified by age group (6 to 12, 13 to 17), by race (White, Black or African American, Other), by sex, by history of prior stimulant medication (yes, no) and by region (Europe, US) will be presented for each of the analysis sets.

The above table will also be provided for Study Part B, using the open-label safety set and Study Part B baseline variables, as applicable.

A by-subject listing will be provided, presenting demographic and other baseline characteristics for subjects in the double-blind safety set.

6.3.2 Medical History and Concurrent Medical Conditions

Summaries of prior medical history and concurrent medical conditions, as collected at Screening (Visit 1A), will be presented by treatment arm and based on the double-blind safety set. No inferential statistics will be presented.

Prior medical history will consist of any significant conditions or diseases that stopped at or prior to time of informed consent. Ongoing conditions will be considered concurrent medical conditions. Concurrent medical conditions are significant ongoing conditions or diseases present at time of informed consent through end of study. These include clinically significant laboratory, ECG, or physical examination abnormalities.

Prior medical history and concurrent medical conditions will be coded using the Medical Dictionary for Drug Regulatory Activities (MedDRA), Version 22.0 or later. The number and percentage of subjects with prior medical history and concurrent medical conditions will be reported by system organ class and preferred term, in separate tables.

A by-subject listing will be presented for medical history, including ongoing conditions, for subjects in the double-blind safety set.

Psychiatric history, as collected in the Kiddie-Schedule for Affective Disorders-Present and Lifetime Version (K-SADS-PL), will be summarized as well, using the double-blind safety set. This table will include time since first onset of ADHD symptoms, time since ADHD diagnosis, ADHD presentation, as well as psychiatric diagnoses and treatments.

Time since first ADHD symptoms/initial ADHD diagnosis will be calculated as per the formula below, rounded to the nearest integer:

- Time since First ADHD Symptom (years): *(Date of Informed Consent – Date of first ADHD Symptom)/365.25*
- Time since Initial ADHD Diagnosis (years): *(Date of Informed Consent – Date of initial ADHD diagnosis)/365.25*

A by-subject listing will be provided, using the double-blind safety set, which presents the ADHD diagnosis history, including start dates for other psychiatric diagnoses.

6.4 Medication History and Concomitant Medications

For Study Part A, any non-study medication that was stopped prior to the first blinded study drug dose date (Day 1A) will be considered prior medication. Any non-study medication with a start date between Day 1A and on or prior to the last blinded study drug dose date, inclusive, or with a start date before Day 1A and an end date after Day 1A or ongoing, will be considered concomitant medication for Study Part A. Any medication with a start date after the last blinded study drug dose date will be considered post-treatment medication for Study Part A. For summary purposes, if the medication's start or stop dates are missing or partially missing and the medication is not checked as ongoing, the following imputation rules will be applied:

- If the year is present but the month and day are missing, then 01JAN will be imputed for the start date and 31DEC for the stop date.
- If the year and month are present but the day is missing, then 01 will be imputed for the start date and the last day of the month for the stop date.
- If the year is missing or the date otherwise cannot be imputed, then the date will be treated as missing and the medication will be treated as a concomitant medication.

All the medication history, concomitant medications, and post-treatment medications will be summarized for the double-blind safety set in frequency tabulations (subject counts and percentages) and by ATC first level and Preferred Name, as coded by World Health Organization Drug Dictionary (WHODrug), Mar 2019 version or later. In Drug Class and Preferred Name summarizations, a subject will be counted once if the subject reports one or more medications.

Analogous classification and presentations will be presented for non-study medications, relative to the first open-label study drug dose date in Study Part B (ie Day 1B), for subjects in the open-label safety set.

A listing of non-study medication will also be provided, to include dose, unit, frequency, route of administration, start and end dates, reason for use, and prior/concomitant/post-treatment flags (for both double-blind and open-label study parts) for subjects in the double-blind safety set.

Procedures and therapies will similarly be classified as prior/concomitant/post-treatment relative to double-blinded dosing, for subjects in the double-blind safety set. Summary tables will be presented summarizing procedures and therapies as prior/concomitant/post-treatment relative to double-blinded dosing, using the double-blind safety set. This will be repeated for Study Part B, using subjects in the open-label safety set. Listings of non-study procedures as well as therapies will also be provided, with prior/concomitant/post-treatment flags (for both double-blind and open-label study parts) for subjects in the double-blind safety set.

In addition, a summary table and listing will be presented for prior stimulant medications, including reason for stopping medication, for subjects in the double-blind safety set.

6.5 Efficacy Analysis

6.5.1 Primary Endpoint(s) Analysis

Not applicable.

6.5.2 Secondary Endpoint(s) Analysis

The analysis set for efficacy summaries and analyses corresponding to Study Part A will be the FAS, while the analysis set for efficacy summaries corresponding to Study Part B will be the open-label safety set.

6.5.2.1 ADHD-RS-5

The ADHD-RS-5 (DuPaul et al., 2016) is used widely by mental health, educational, and medical practitioners in screening, diagnosis, and treatment evaluation to determine the frequency and severity of ADHD symptoms and impairments in children and adolescents. The ADHD-RS-5 is based on the diagnostic criteria for ADHD as described in the DSM-5 and consists of 2 symptom subscales, inattention (Questions 1 to 9) and hyperactivity-impulsivity, (Questions 16 to 24), and a total scale of 18 items.

Each item in the subscale is scored with a value ranging from 0 (no symptoms) to 3 (severe symptoms). The subscale scores can range from 0 to 27, while the total score can range from 0 to 54, with a higher score indicating worse outcome. For each subscale, if up to 1 of the items is missing for a given visit, then the subscale score will be prorated using the non-missing values (ie multiply the mean of the non-missing items by the total number of items [9] in the subscale.) Similarly, for the total score, if up to 3 items are missing, then the total score may be prorated accordingly. Otherwise, if ≥20% of items are missing, then the corresponding score will be set to missing.

Observed values, as well as absolute and percent change from baseline, for the subscale and total scores, will be summarized descriptively, for each visit in Study Part A. This table will also be presented summarized by age group (6 to 12, 13 to 17), by race (White, Black or African American, Other), by history of prior stimulant medication (yes, no), as well as by sex. In addition, this table will be repeated, stratified by region (Europe, US), to account for potential COVID-19 related differences. Additional subgroups may be identified, as the data may indicate, and summarized as described above.

The change from baseline in ADHD-RS-5 subscale and total scores for each scheduled post-baseline visit in Study Part A (including scheduled Part A visits for which data were collected prior to those visits being eliminated under Protocol Amendment 7) will be evaluated via MMRM, as described in Section 6.1.2.

In addition, the proportion of responders (defined as a reduction from baseline ADHD-RS-5 total score of ≥30% and CGI-I of 1 or 2 at a given post-baseline visit) in each treatment arm and the difference in the proportion of responders between treatment arms will be summarized for each

scheduled post-baseline visit in Study Part A (including scheduled Part A visits for which data were collected prior to those visits being eliminated under Protocol Amendment 7), along with the corresponding 95% confidence intervals, as described in Section 6.1.3.

The following 6 domains of impairment common among children and adolescents with ADHD are also assessed with the ADHD-RS-5:

- Relationships with significant others (such as family members and teachers)
- Relationships with peers
- Academic functioning
- Behavioral functioning
- Homework performance
- Self-esteem

For each subscale, the 6 domains will be summarized descriptively, stratified by age group (6 to 12, 13 to 17), for each visit in Study Part A.

Figures will be presented displaying observed mean change from baseline values (with 95% CI bars) by treatment arm, for the subscales and total scores, for each scheduled post-baseline visit in Study Part A (including scheduled Part A visits for which data were collected prior to those visits being eliminated under Protocol Amendment 7).

All tables and figures described above (excluding MMRM and CMH analyses) will be repeated for Study Part B, using the open-label safety set.

In addition, in order to present the change in ADHD-RS-5 during the entire course of the study, a figure will be presented displaying observed mean change from baseline values (with 95% CI bars) across Part A and B, by treatment arm in Part A, for each scheduled post-baseline visit. The baseline will be defined as the last observed value before the first dose of blinded IMP.

A by-subject listing will be provided, which presents ADHD-RS-5 version and values (for item, subscale and total scores) across all assessments in both study parts, for subjects in the FAS.

6.5.2.2 CGI-I

The CGI scale (Guy, 1976) will be used to evaluate the severity of mental illness over time. The CGI-S will be administered to assess the severity of mental illness at baseline. During subsequent visits, the CGI-I will be administered to assess any improvement in symptoms and to guide the clinician on dosing adjustments. The CGI-S is scored on a 7-point scale ranging from 1 (normal, not at all ill) to 7 (among the most extremely ill subjects). The CGI-I is also scored on a 7-point scale ranging from 1 (very much improved) to 7 (very much worse).

A frequency table will display the distribution of CGI scores in each treatment arm for each visit of Study Part A. In addition, the percentage of subjects who had a score on the CGI-I indicating improvement (defined as "very much improved" or "much improved") will be presented for each post-baseline visit, along with the differences (and 95% CI) between treatment arms. This table will also be presented summarized by age group (6 to 12, 13 to 17), by race (White, Black or African American, Other), by history of prior stimulant medication (yes, no), as well as by sex. In addition, the above table will be repeated for Study Part A, stratified by region (Europe, US), to account for potential COVID-19 related differences.

Cochran Mantel Haenzsel tests will be performed to evaluate differences between the proportions of subjects who had a score indicating improvement for each scheduled post-baseline visit in Study Part A (including scheduled Part A visits for which data were collected prior to those visits being eliminated under Protocol Amendment 7), as described in section 6.1.3.

A figure will be presented which displays the proportion of subjects in each treatment arm with CGI-I improvement, along with 95% CI, for each scheduled post-baseline visit in Study Part A (including scheduled Part A visits for which data were collected prior to those visits being eliminated under Protocol Amendment 7).

All tables and figures described above (excluding the CMH analysis) will be repeated for Study Part B, using the open-label safety set.

A by-subject listing will be provided, which presents CGI-S and CGI-I scores across all assessments in both study parts, for subjects in the FAS.

6.5.2.3 CHIP-CE-PRF

The Parent Report Form of the Child Health and Illness Profile – Child Edition (CHIP-CE:PRF) will be administered to provide information on self–esteem and school functioning in pediatric subjects diagnosed with ADHD. As reported by a parent or caregiver, a child's functioning and well-being as well as emotional, physical, and behavioral symptoms will be captured with this generic scale that is useful for monitoring symptoms and identifying improvements. The 5 domains and 12 subdomains covered in the 76 items comprising the CHIP-CE:PRF are presented below in Table 2. Each item uses a 5-response format (scored 1 to 5), with a higher score indicating greater health. Accordingly, several items will be reverse scored, as indicated in the technical manual for the parent report form (Riley, 2001).

Table 2 Domains and Subdomains of the CHIP – CE:PRF

Domain	Subdomain
Satisfaction: with health and self	Satisfaction with health (7 items-1,2,3,4,10,11,12)
	Satisfaction with self (4 items-6,7,8,9)

Table 2 Domains and Subdomains of the CHIP – CE:PRF

Domain	Subdomain
Comfort: physical and emotional symptoms and activity restrictions due to illness	Physical comfort (9 items- 5, 13, 14, 15, 16, 17, 18, 19 20)
	Emotional comfort (9 items- 21, 22, 23, 24, 25, 26, 27, 28, 29)
	Restricted activity (4 items- 30, 34, 35, 36)
Resilience: behaviors and family involvement in activities likely to enhance health	Family involvement (8 items- 37, 38, 39, 40, 41, 42, 43, 47)
	Social problem-solving (5 items- 64, 65, 66, 67, 68)
	Physical activity (6 items- 31, 32, 33, 44, 45, 46)
Risk avoidance: behaviors that if not avoided are likely to pose risks to health	Individual risk avoidance (4 items- 48, 49. 50, 51)
	Threats to achievement (11 items- 56, 57, 58, 59, 60, 61, 62, 63, 73, 74, 76)
Achievement: developmentally appropriate role functioning in school and with peers	Academic performance (5 items- 69, 70, 71, 72, 75)
	Peer relations (4 items- 52, 53, 54, 55)

CHIP-CE:PRF=Child Health and Illness Profile-Child Edition: Parent Report Form

For each domain/subdomain, means will be calculated by taking the average of each non-missing item in the corresponding domain/subdomain (where <30% of items are missing). The global score is calculated by taking the average score across the five domains (no domains can be missing). These values, as well as absolute and percent change from baseline, will be summarized descriptively, for each visit in Study Part A. This table will also be presented summarized by age group (6 to 12, 13 to 17), by race (White, Black or African American, Other), by history of prior stimulant medication (yes, no), as well as by sex. In addition, this table will be repeated, stratified by region (Europe, US), to account for potential COVID-19 related differences.

The change from baseline in the CHIP-CE:PRF global, domain and subdomain scores for each scheduled post-baseline visit in Study Part A (including scheduled Part A visits for which data were collected prior to those visits being eliminated under Protocol Amendment 7) will be evaluated via MMRM, as described in Section 6.1.2.

Figures will be presented displaying observed mean change from baseline values (with 95% CI bars) by treatment arm, for the global and domain scores, for each scheduled post-baseline visit in Study Part A (including scheduled Part A visits for which data were collected prior to those visits being eliminated under Protocol Amendment 7).

All tables and figures described above (excluding the MMRM analysis) will be repeated for Study Part B, using the open-label safety set.

A by-subject listing will be provided, which presents all CHIP-CE:PRF observed values, along with global, domain and subdomain scores, across all assessments in both study parts, for subjects in the FAS.

6.5.2.4 Conners 3 Parent Short Form

Subjects enrolled after Amendment 5 will be assessed with the C3PS (Conners, 2008). This assessment is used to measure behavioral, social, and academic issues in children between 6 and 18 years old. The scale aids in diagnosis by helping to delineate the child's issues, as well as in what settings these issues are most troublesome.

The scale consists of 45 items, 43 of which are scored 0-3. Total scores will be evaluated by summing the 43 numeric items, with a higher score indicating greater frequency of issues. If up to 8 items are missing for a given assessment, then the total score may be prorated accordingly (ie multiply the mean of the non-missing items by the total number of items [43] in the scale). Otherwise, if ≥20% of items are missing for a given assessment, then the total score will be set to missing. The C3PS has the following subscales: Inattention, Hyperactivity/Impulsivity, Learning Problems, Executive Functioning, Defiance/Aggression, Peer Relation and Validity Scores (Positive Impression and Negative Impression). Table 3 below shows the items which correspond to the respective subscales and validity scales of the C3PS.

Table 3 Subscales and Validity Scales of the C3PS

Subscale	Items
Inattention	(5 items) 17. Doesn't pay attention to details; makes careless mistakes. 27. Has a short attention span. 30. Has trouble concentrating. 34. Inattentive, easily distracted. 41. Has trouble keeping his/her mind on work or play for long.
Hyperactivity/Impulsivity	(6 items) 3. Fidgets or squirms in seat. 5. Restless or overactive. 7. Runs or climbs when he/she is not supposed to. 13. Acts as if driven by a motor. 24. Is constantly moving 28. Excitable, impulsive.
Learning Problems	(5 items) 8. Cannot grasp arithmetic. 10. Needs extra explanation of instructions. 25. Has trouble with reading. 36. Spelling is poor. 39. Does not understand what he/she reads.
Executive Functioning	(5 items) 1. Forgets to turn in completed work. 15. Has trouble getting started on tasks or projects. 20. Loses things (for example, schoolwork, pencils, books, tools, or toys). 32. Has trouble organizing tasks or activities. 35. Is messy or disorganized.
Defiance/Aggression	(5 items) 14. Starts fights with others on purpose. 19. Bullies, threatens, or scares others. 21. Tells lies to hurt other people. 23. Threatens to hurt others. 26. Is angry and resentful.
Peer Relations	(5 items) 4. Is one of the last to be picked for teams or games. 6. Does not know how to make friends. 18. Has trouble keeping friends. 38. Has no friends. 43. Does not get invited to play or go out with others.

Subscale	Items
Validity Scales	
Positive Impression	(6 items) 2. Is perfect in every way. 12. Makes mistakes. (R) 31. Tells the truth; doesn't even tell "little white lies." 37. Is patient and content, even when waiting in a long line. 40. Behaves like an angel. 42. Has to struggle to complete hard tasks. (R)
Negative Impression	(6 items) 9. Is difficult to please or amuse. 11. Is hard to motivate (even with rewards like candy or money). 16. Is happy, cheerful, and has a positive attitude. (R) 22. I cannot figure out what makes him/her happy. 29. Cannot do things right. 33. Is fun to be around. (R)

Among these subscales, all of the items included in the Learning Problems and Executive Functioning subscales either directly or indirectly evaluate performance related to schoolwork. Responses for these 2 subscales will be converted to T-scores, following standardization against age and gender specific normative data of a North American general reference population.

Observed values, as well as absolute and percent change from baseline, for the total score and the two aforementioned subscale T-scores, will be summarized descriptively, for each visit in Study Part A. This table will also be presented summarized by age group (6 to 12, 13 to 17), by race (White, Black or African American, Other), by history of prior stimulant medication (yes, no), as well as by sex. In addition, this table will be repeated, stratified by region (Europe, US), to account for potential COVID-19 related differences.

The change from baseline in the C3PS total score for each scheduled post-baseline visit in Study Part A (including scheduled Part A visits for which data were collected prior to those visits being eliminated under Protocol Amendment 7) will be evaluated via MMRM, as outlined in Section 6.1.2.

Figures will be presented displaying observed mean change from baseline values (with 95% CI bars) by treatment arm, for the total and the two aforementioned subscale T-scores, for each scheduled post-baseline visit in Study Part A (including scheduled Part A visits for which data were collected prior to those visits being eliminated under Protocol Amendment 7).

All tables and figures described above (excluding the MMRM analysis) will be repeated for Study Part B, using the open-label safety set.

A by-subject listing will be provided, which presents all C3PS observed values, along with total and subscale T-scores, across all assessments in both study parts, for subjects in the FAS.

6.5.2.5 Sensitivity Analysis

- To assess the treatment impact without influence of the COVID-19 pandemic, a sensitivity analysis will be performed, if data permit, in an analogous manner to the one described in section 6.6.1.3, for each secondary efficacy endpoint, identified above.
- The study team became aware of data collection deviations (ie assessments administered via paper rather than electronically, C-SSRS raters not having administrative registration or full verification of certification) at individual subject levels for several assessments that were administered via the CamCog device (i.e. ADHD-RS-5, CGI-I, BPRS-C, C-SSRS, UKU). The number of impacted observations is limited, and additionally, the team has concluded that these deviations do not serve to impact the validity of the collected data. As such, presently, this SAP does not propose any adjustments to the planned analyses. The overall scope and impact of this issue is being fully investigated and will be explained in the clinical study report. However, if the investigations suggest adjustments to the analyses, including sensitivity analysis (eg analysis with and without affected data), then these will be delineated prior to database lock. Simple sensitivity analyses may be performed without an amendment to the SAP; however, analyses that go beyond being considered "sensitivity", (eg changes to the preplanned primary analysis) will be documented in an SAP amendment, if needed. Observations affected by these deviations are captured via the protocol deviations list.

6.5.3 Other Secondary Endpoints Analysis

The APRS assessment was removed as a secondary efficacy endpoint with Protocol Amendment 5, due in large part to the COVID-19 pandemic, since many subjects enrolled prior to Amendment 5 are missing APRS data. All APRS data that have been collected will be summarized descriptively (for total and subscale scores) across visits (including unscheduled visits) in both Study Part A and B, using the FAS and open-label safety set, respectively. Visits for the APRS will be presented as collected on the CRF (i.e. not derived via analysis visit windowing).

The three subscales are academic success (Questions 3, 4, 5, 8, 10, 11, 17), impulse control (Questions 9, 12, 16), and academic productivity (Questions 1 to 7, 13, 14, 15, 18, 19). To compute the subscale score, the rating given to each corresponding item on the subscale will be summed. Ratings for items 12, 13, 15, 16, 17, 18, and 19 will be reverse-scored so that high ratings correspond with positive academic functioning.

A by-subject listing will be provided, which presents all observed values, along subscale scores, across all assessments in both study parts, for subjects in the FAS.

Since the APRS waiver was implemented at a distinct timepoint, impacting all subjects equally and simultaneously (eg systematic bias), subsequent missing APRS assessments will be assumed

to be missing at random. No imputation will be performed for missing APRS assessments. Unless data permit, no inferential analysis of this endpoint will be performed. Missing APRS assessments will not be a reason for exclusion from any analysis set.

6.5.4 Subgroup Analyses

In addition to the geographic subgrouping of summary tables described above for secondary efficacy endpoints, subgroup analysis will be performed and presented, in an analogous manner to the one described in section 6.6.1.4, for all secondary efficacy endpoints for which an MMRM is used for analysis. For the CGI-I and ADHD-RS-5 responders endpoints, the CMH analysis will be repeated, with region (Europe or US) added as a stratification factor.

6.6 Safety Analysis

For Study Part A, all safety analyses will use the double-blind safety set, except for analyses of the CANTAB domains, which will use the FAS (and PPS for sensitivity analyses). For Study Part B, all safety presentations will use the open-label safety set.

6.6.1 Primary Safety Endpoint

6.6.1.1 Derivation of Endpoint(s)

In the CANTAB RTI task, a yellow dot will appear in one of the circles (one circle for the simple variant, and five for the five-choice variant) and the participant must react as soon as possible, releasing the button at the bottom of the screen, and selecting the circle in which the dot appeared. The time taken from the yellow dot appearing, to the subject releasing the press-pad is defined as the reaction time. The movement time is defined as the time taken from the subject releasing the press-pad to touching the screen. Times for this assessment are calculated for correct trials, and measured in msec ranging from 100 to 5100, with a higher time indicating worse performance of the task.

6.6.1.2 Main Analytical Approach

Observed values for mean reaction time and mean movement time for both the simple and five-choice variants, as well as absolute and percent change from baseline, will be summarized descriptively, for each visit in Study Part A (including Part A visits for which data were collected prior to those visits being eliminated under Protocol Amendment 7). The time obtained prior to first dose of blinded IMP will serve as each subject's baseline value. This table will also be presented summarized by age group (6 to 12, 13 to 17), by race (White, Black or African American, Other), by history of prior stimulant medication (yes, no), as well as by sex. In addition, this table will be repeated, stratified by region (Europe, US), to account for potential COVID-19 related differences.

The primary safety endpoint is the change from baseline in reaction time (for correct trials) for the Five-Choice CANTAB RTI task, and will be based on the FAS population for visits in Study Part A. As suggested by the European Medicines Agency (EMA), using data up to 18 weeks, the

superiority to placebo comparisons of TAK-503 and atomoxetine, separately, will be performed using the MMRM as specified for the primary endpoint. If TAK-503 is statistically superior to placebo, then a negative impact on cognition can be excluded. If atomoxetine is statistically superior to placebo, then the assay sensitivity is demonstrated.

If neither TAK-503 nor atomoxetine is superior to placebo, then the noninferiority between TAK-503 and placebo at 18 weeks will be assessed using the margin of mean change from baseline in placebo arm. The 1-sided 97.5% confidence interval (upper bound) for the difference between TAK-503 and placebo will be derived from the MMRM. If the upper bound of the 1-sided 97.5% CI for the difference lies entirely below the margin (in msec), then it will be concluded that TAK-503 is noninferior to placebo for the primary endpoint.

In addition, the effect of TAK-503 and atomoxetine will be compared at 1 year (Week 49A). Prior to elimination of the 18 to 49 week portion of Study Part A in Protocol Amendment 7, data for CANTAB domains (and other assessments) were collected through 49 weeks in Study Part A for active arm subjects.

The analysis of the primary endpoint, as well as other variables within CANTAB RTI, will be conducted via MMRM, as described in Section 6.1.2. The corresponding statistical hypotheses are as defined in Section 3.1. The strategy for handling intercurrent events (such as those related to treatment discontinuation or COVID-19 impact) will follow the estimand framework outlined in Section 1.3.

Where applicable, descriptive statistics, including 95% CIs, will be provided for observed values and change from baseline values for Study Part B, using the open-label safety set.

Figures will be presented, displaying observed mean change from baseline values (with 95% CI bars) for both the reaction and movement times in each variant (simple and five-choice) of the CANTAB RTI Task by treatment arm for each scheduled post-baseline visit in Study Part A (including scheduled Part A visits for which data were collected prior to those visits being eliminated under Protocol Amendment 7), using subjects in the FAS. Similar figures will be provided for the scheduled visits of Study Part B, using the open-label safety set. Besides, a figure will be presented displaying observed mean change from baseline values (with 95% CI bars) across Part A and B, by treatment arm in Part A, for each scheduled post-baseline visit. The baseline will be defined as the last observed value before the first dose of blinded IMP.

In addition, a forest plot will be provided, showing the difference in LS mean estimates (with 95% CI) for the change from baseline to Week 18A between each active arm and placebo, for each CANTAB domain, using the FAS. A similar plot will be provided, comparing the estimated change from baseline to Week 18A and Week 49A between TAK-503 and atomoxetine.

A by-subject listing will be provided, which presents results for both the simple and five-choice CANTAB RTI task across all assessments in both study parts, for subjects in the FAS.

6.6.1.3 Sensitivity Analysis

To investigate robustness of the results, the below procedures will be conducted as sensitivity analyses of the primary safety endpoint. These analyses will be performed in a similar manner for additional endpoints, where specified:

- The MMRM analysis, as described in 6.1.2, will be repeated for scheduled visits of Study Part A using the PPS. For comparisons of CANTAB domains through Week 18A which use the PPS, only PPS subjects who complete Week 18A (for all treatment arms) will be included in the MMRM. For the Week 49A comparison, only PPS subjects who complete Week 49A will be included in the MMRM (ie will exclude placebo subjects and subjects enrolled under Protocol Amendment 7 or later). This sensitivity analysis will be repeated for all other CANTAB domains, as well.
- If data permit, to estimate the treatment impact without influence of the COVID-19 pandemic, a sensitivity analysis will be run which excludes subjects who had a COVID-19 Impact CRF populated for any time during Study Part A. The MMRM, as described in section 6.1.2, will be run excluding subjects who have any visit or assessment missed/impacted by COVID-19 during Study Part A. This sensitivity analysis will be repeated for all other CANTAB domains and secondary efficacy endpoints, as well.
- If data permit, to estimate the impact of prior stimulant treatment, a sensitivity analysis will be run for CANTAB RTI which excludes the site where the most subjects without prior stimulant treatment are from.
- If data permit, to estimate the impact of alternative methods of contact for data collection, the MMRM used for the primary model will be repeated with method of contact (alternative vs. standard) added as a term in the model, along with treatment*method of contact. This sensitivity analysis will be repeated for all other CANTAB domains and secondary efficacy endpoints for Study Part A, as well.

6.6.1.4 Subgroup Analyses

To further examine regional differences in treatment effect, the MMRM, using subjects from the FAS, will be repeated for the primary endpoint (mean reaction time and mean movement time for both the simple and five-choice variant), with the addition of an interaction term for treatment*region. That is, the model will include treatment arm, visit, sex (male or female), age group (6 to 12 years or 13 to 17 years), region (Europe or US), the interaction of treatment arm with visit, and the interaction of treatment arm with region as factors, the corresponding baseline value as a covariate, and the interaction of baseline value with visit adjusted in the model. The LS mean estimates of the change from baseline to each scheduled post-baseline visit in Study Part A (including scheduled Part A visits for which data were collected prior to those visits being eliminated under Protocol Amendment 7) for each treatment arm, the differences between the LS means across treatment arms, and 95% confidence intervals and p-values associated with the differences will be reported at each visit, as applicable. In addition, the (nominal) p-value (Type III) for the treatment*region interaction term will be reported for each endpoint.

A similar subgroup analysis will be performed using age group (6 to 12, 13 to 17) as the subgroup of interest to examine differences amongst the child and adolescent populations.

The above subgroup analyses will be repeated for the other CANTAB domains and secondary efficacy endpoints for Study Part A using the FAS, as well.

6.6.2 Secondary Safety Endpoints

6.6.2.1 CANTAB: Rapid Visual Processing (RVP)

The RVP task measures the ability to sustain attention over time and is a sensitive measure of frontal-parietal function. Subjects are to detect a 3-digit target sequence (e.g., 2-4-6) and respond by pressing a button at the bottom of the screen when the final number of the sequence appears on the screen.

The outcome measures to be presented for each assessment of this task are as follows:

- A': standardized score for target sequence detection, ranging from 0-1 (higher score indicates better performance)
- Mean Response Latency: the mean response time on trials where subject responded correctly, measured in msec (higher time indicates worse performance)
- Probability of Hit: Proportion of correct sequence responses divided by total number of sequences (higher rate indicates better performance)

These outcome measures will be summarized and analyzed similarly to the outcome measures of the primary safety endpoint (CANTAB RTI task), as outlined in Sections 6.6.1.2-6.6.1.4.

6.6.2.2 CANTAB: Spatial Working Memory (SWM)

The SWM task measures the ability to retain spatial information and manipulate remembered items in working memory are measured with the SWM task. The task is self-ordered and assesses the individual's ability to strategize heuristically. The test begins with a number of colored squares (boxes) shown on the screen. By selecting the boxes and using a process of elimination, the participant should find one yellow 'token' in each of a number of boxes and use them to fill up an empty column on the right-hand side of the screen.

The outcome measures to be presented for each assessment of this task are as follows:

- Strategy (6 to 8 boxes): Strategy score based on the number of time a subject begins a new search pattern from the same box they started with previously (higher score indicates worse performance)
- Total Errors: The number of times a box was selected that was certain to not have any tokens, across all trials (higher value indicates worse performance)

These outcome measures will be summarized and analyzed similarly to the outcome measures of the primary safety endpoint (CANTAB RTI task), as outlined in Sections 6.6.1.2-6.6.1.4.

6.6.2.3 CANTAB: Stop Signal Task (SST)

The SST measures response inhibition. The subject must respond to an arrow stimulus by touching either of 2 choices depending on the direction the arrow points. If an audio tone is present, the subject is not to respond.

The outcome measures to be presented for each assessment of this task are as follows:

- Stop Signal Reaction Time: The estimate of time where an individual can successfully inhibit responses 50% of the time (higher time indicates better performance)
- Direction Error (Go Trials): The total number of trials where the subject pressed the wrong button to the direction of the arrow stimulus on a "Go" trial (higher value indicates worse performance)
- Direction Error (Stop Trials): The total number of trials where the subject pressed the wrong button to the direction of the arrow stimulus on a "Stop" trial (higher value indicates worse performance)
- Median Reaction Time (All Go Trials): Median reaction time taken across all Go trials within an assessment (higher time indicates worse performance)
- Missed Trials: The total number of trials which were missed by the subject (higher value indicates worse performance)

These outcome measures will be summarized similarly to the outcome measures of the primary safety endpoint (CANTAB RTI task), as outlined in Section 6.6.1.2-6.6.1.4.

6.6.2.4 CANTAB: Delayed Matching to Sample (DMS)

The DMS tests both simultaneous matching and short-term visual memory. The subject is shown a complex visual pattern (the sample) and after a brief delay, 4 similar patterns. The subject must identify the pattern that matches the sample.

The outcome measures to be presented for each assessment of this task are as follows:

- Mean Choices to Correct: the mean number of choices that the subject made on each trial, including the correct choice (higher number of choices indicates worse performance)
- Mean Correct Latency: the average time between the presentation of the response stimuli objects and the subject selecting the correct box on their first attempt (higher time indicates worse performance)
- Proportion of Correct Responses: the percentage of trials during which the subject chose the correct response on the first attempt (higher rate indicates better performance)

These outcome measures will be summarized and analyzed similarly to the outcome measures of the primary safety endpoint (CANTAB RTI task), as outlined in Section 6.6.1.2-6.6.1.4.

6.6.2.5 Tanner Staging

The stage of puberty/sexual maturation will be evaluated for each subject according to Tanner staging (Marshall and Tanner, 1969; Marshall and Tanner, 1970). The Tanner stage for genitals (male, stages I to V), breasts (females, stages I to V), and pubic hair (both sexes, stages I to V) will be documented at the times specified in the Protocol.

Tanner staging will be self-assessed. Self-assessment in this study is defined as subjects or parents indicating which drawing of the scale corresponds to subject's sexual maturation stage at the time of the specific visit. For this purpose, site staff (the principal investigator or a designee) reads the guidance and the text corresponding to each drawing and asks the subject or parent to choose the applicable drawing (see Appendix 7 in Protocol). The response from the subject or the parent will be documented in the Tanner Staging Form by the site staff.
Tanner stages will be summarized descriptively by treatment arm, for visits in Study Part A for subjects in the double-blind safety set, stratified by sex and age group (6 to 12, 13 to 17). A shift table from baseline to each visit will also be presented, stratified by sex and age group (6 to 12, 13 to 17).

In addition, age will be summarized descriptively by treatment arm for each visit in Study Part A for each Tanner stage, stratified by sex.

The above tables will be repeated for visits in Study Part B, using the open-label safety set.

A by-subject listing will present Tanner stages across all assessments in both study parts, for subjects in the double-blind safety set.

6.6.2.6 Brief Psychiatric Rating Scale for Children (BPRS-C)

The BPRS-C-21 was developed to provide a concise symptom profile of psychiatric problems that can occur during childhood (Hughes 2008, Hughes et al., 2001; Overall and Pfefferbaum, 1982). Each of the 21 items are rated on a 7-point severity Likert scale, ranging from 0 (not present) to 6 (extremely severe symptom). A total score is computed (where <20% of items are missing) by summing the values across the 21 items, thus ranging from 0 to 126. If fewer than 5 items are missing, then the total score can be derived by prorating the values for the non-missing items.

The 21 questions are grouped across 7 scales (Behavior Problems [Questions 1 to 3], Depression [Questions 4 to 6], Thinking Disturbance [Questions 7 to 9], Psychomotor Excitation [Questions 10 to 12], Withdrawal [Questions 13 to 15], Anxiety [Questions 16 to 18] and Organicity [Questions 19 to 21]), and for each scale, a score will be computed by summing the values across the 3 items (set to missing if any items are missing) in the corresponding scale.

The total and scale scores, along with absolute and percent change from baseline, will be summarized descriptively by treatment arm for visits in Study Part A for subjects in the double-blind safety set. This table will be repeated, stratified by age group (6 to 12, 13 to 17) and another table stratified by sex.

The change from baseline in total score for each scheduled post-baseline visit in Study Part A (including scheduled Part A visits for which data were collected prior to those visits being eliminated under Protocol Amendment 7) will also be analyzed via MMRM, as outlined in Section 6.1.2, using the double-blind safety set.

Subjects enrolled prior to Amendment 5 were administered a version of this scale which omitted the final 8 items. For subjects administered a 13 item scale at baseline (for Study Part A or B), then the total score and scores for the withdrawal, anxiety and organicity scales will not be presented in the tables for the corresponding post-baseline visits, even if the subject was administered the 21 item assessment at a post-baseline visit. All collected data will be displayed in the listings. A sensitivity analysis to impute scores for assessments using the 13 items may be considered.

A figure will be presented, displaying observed mean values of the total score (with 95% CI bars) by treatment arm for each scheduled visit in Study Part A (including scheduled Part A visits for which data were collected prior to those visits being eliminated under Protocol Amendment 7).

All tables and figures described above (except the MMRM analysis) will be repeated for Study Part B, using the open-label safety set.

A by-subject listing will be provided, which presents BPRS-C assessment values across all assessments in both study parts, for subjects in the double-blind safety set.

6.6.2.7 Columbia Suicide Severity Rating Scale (C-SSRS)

Columbia Suicide Severity Rating Scale is a suicide assessment questionnaire, which evaluates suicidal ideation and behavior. A maximum of 19 items will be completed as follows: 7 items are required, a potential 10 additional items will be completed upon a positive response to a required item, and 2 items completed if suicide/suicide-like behavior is observed during the interview.

For study Part A, descriptive statistics will be presented as follows, using the double-blind safety set, presented by visit and overall (post-baseline):

- Number of subjects with any suicidal ideation, any suicidal behavior, and any suicidal ideation or behavior at each visit. Number of subjects with self-injurious behavior without suicidal intent will also be presented
- Number of subjects who answered yes for each of the “yes/no” questions
- Number of actual/interrupted/aborted attempts, along with actual and potential lethality categories for actual attempts

In addition, these tables will be repeated, stratified by region (Europe, US), to account for potential COVID-19 related differences.

Visits for the C-SSRS will be presented as collected on the CRF (ie not derived via analysis visit windowing).

All tables described above will be repeated for Study Part B, using the open-label safety set.

A by-subject listing will be provided, which presents responses to the C-SSRS questionnaire across all assessments in both study parts, for subjects in the double-blind safety set. A separate listing will be provided, displaying only subjects who answered yes to any of the "yes/no" questions.

6.6.2.8 Udvalg for Kliniske Undersøgelser (UKU)

Designed for use in both clinical trials and routine clinical practice, the UKU rating scale was developed for clinicians to assess side effects of psychopharmacological medications based on patient interviews and other relevant source information (Lingjaerde et al., 1987). For this study, only the following items relevant to the established safety profile of TAK-503 will be queried: Asthenia/Lassitude/lncreased Fatiguability, Sleepiness/Sedation, Increased Duration of Sleep, and Orthostatic Dizziness. Each side effect is categorized for severity, ranging from 0 (Normal) to 3 (Severe). In addition, each side effect's relationship to the administered drug is assessed, and categorized as "Impossible," "Possible," and "Probable."

Each item will be summarized descriptively by treatment arm, with categories presented for severity and causal relationship, along with the global assessments of side effects' interference, for visits in Study Part A for subjects in the double-blind safety set. Visits for the UKU will be presented as collected on the CRF (i.e. not derived via analysis visit windowing).

The above will be repeated for visits in Study Part B for subjects in the open-label safety set.

A by-subject listing will be provided, which presents responses to the UKU rating scale, including severity rating and causal relationship, across all assessments in both study parts, for subjects in the double-blind safety set.

6.6.2.9 Pediatric Daytime Sleepiness Scale (PDSS)

The PDSS is a self-reported assessment of daytime sleepiness in children **(**Drake et al., 2003). The 8 questions are scored on Likert-scale from 0 to 4 (never=0; seldom=1; sometimes=2; frequently=3; always=4). The total score on the PDSS can be derived by summing the values from the 8 questions (where <20% of the questions are missing), and thus range from 0 (never sleepy) to 32 (always sleepy). In order to reduce the possibility of response bias, the response to item number 3 is reverse scored. If 1 of the questions has a missing value, then the total score can be derived by prorating the values for the non-missing questions. Otherwise, if more than 1 of the questions has a missing value, then the total score will be set to missing.

The total scores, along with absolute and percent change from baseline, will be summarized descriptively by treatment arm, for visits in Study Part A for subjects in the double-blind safety set. This table will also be presented summarized by age group (6 to 12, 13 to 17) as well as by sex.

The change from baseline in total score for each scheduled post-baseline visit in Study Part A (including scheduled Part A visits for which data were collected prior to those visits being eliminated under Protocol Amendment 7) will also be analyzed via MMRM, as outlined in Section 6.1.2, using the double-blind safety set.

A figure will be presented, displaying observed mean values of the total score (with 95% CI bars) by treatment arm for each scheduled visit in Study Part A (including scheduled Part A visits for which data were collected prior to those visits being eliminated under Protocol Amendment 7).

All tables and figures described above (except the MMRM analysis) will be repeated for Study Part B, using the open-label safety set.

A by-subject listing will be provided, which presents PDSS assessment values across all assessments in both study parts, for subjects in the double-blind safety set.

6.6.3 Adverse Events

For Study Part A, treatment-emergent adverse events (TEAEs) are defined as AEs whose onset occurs, severity worsens or intensity increases after receiving the blinded IMP and up to 3 days after the last dose of double-blind study medication (including those which may have occurred during Study Part B screening prior to first dose of open-label TAK-503). For Study Part B, treatment-emergent adverse events are defined as AEs whose onset occurs, severity worsens or intensity increases after receiving the open-label TAK-503 and up to 3 days after the last dose of open-label TAK-503. Treatment-related adverse events will be defined as treatment-emergent AEs which were identified by the investigator to be related to study drug.

Adverse event verbatim reported terms will be coded by system organ class (SOC) and preferred term (PT) using MedDRA, Version 22.0 or later. AE summary tables will include numbers and percentages of subjects experiencing at least one AE by SOC and PT, in addition to the numbers of events.

Furthermore, to account for differences in expected exposure durations, time adjusted event rates (per 100 years) will be presented in overall summary tables, as well. For each study part, this rate will be calculated as:

(Number of TEAE's in Study Part)/(Total Person Years at Risk in Study Part)*100

For each subject, the number of person years at risk is calculated as:

(Last Contact Date in Study Part – Date of First Exposure in Study Part +1)/365.25

Unless otherwise specified, AEs will be sorted by relative incidence (highest to lowest) in the TAK-503 arm, unless where the AE summary is two-tiered, then the AE will be sorted alphabetically at the first tier (eg SOC) and relative incidence at the second tier (eg PT). The following AE summary tables will be generated, for each of the double-blind and open-label study parts, using the double-blind and open-label safety sets, respectively:

- Overview of AEs (repeat stratified by region, age group (6 to 12, 13 to 17), race (White, Black or African American, Other), weight-adjusted TAK-503 dose at time of onset (<=0.04 mg/kg; 0.04 mg/kg ~ <=0.08 mg/kg; 0.08 mg/kg ~ <=0.12 mg/kg; >0.12 mg/kg) and actual TAK-503 dose at time of onset (7 subgroups from 1 mg to 7 mg)).
- Pre-treatment AEs by SOC and PT
- TEAEs by SOC and PT (repeat stratified by region (Europe, US), age group (6 to 12,13 to 17), race (White, Black or African American, Other), weight-adjusted TAK-503 dose at time of onset and actual TAK-503 dose at time of onset)
- Common TEAEs (Incidence above a threshold to be determined upon data review) by PT
- Serious TEAEs (TESAE) by SOC and PT
- Sedative TEAE's (PT= Somnolence, Sedation, Hypersomnia, or Fatigue) by SOC and PT (repeat stratified by region (Europe, US), age group (6 to 12, 13 to 17), race (White, Black or African American, Other) weight-adjusted TAK-503 dose at time of onset and actual TAK-503 dose at time of onset)
- Relationship of TEAEs to Study Drug by SOC and PT
- Relationship of TEAEs to Study Procedure by SOC and PT
- Drug-Related TEAEs by SOC and PT
- Procedure-Related TEAE by SOC and PT
- ADHD-Related TEAE by SOC and PT
- Severity of TEAEs by SOC and PT
- TEAE with Action of Dose Interrupted by SOC and PT
- TEAE with Action of Drug Withdrawn SOC and PT
- TEAE leading to Study Discontinuation by SOC and PT
- TEAE leading to death by SOC and PT
- COVID-19 related AE by SOC and PT
- TEAE by SOC and PT for COVID-19 infected subjects

In addition, the following AE tables (with time adjusted event rates) will be provided for subjects in the double-blind safety set, using data spanning both study parts. That is, data will be grouped by Placebo (include data from Study Part A), Atomoxetine (include data from Study Part A), and TAK-503 arms (include data from Study Part A and/or B):

- TEAE by SOC and PT
- Most frequent (>=2%) non-serious TEAE by SOC and PT

The following listings will also be provided, using data from both study parts for subjects in the double-blind safety set:

- All unique MedDRA terms
- Serious Adverse Events
- Adverse Events Leading To Drug Withdrawal
- Adverse Events Leading To Study Discontinuation
- Adverse Events Resulting in Death
- All Adverse Events in COVID-19 infected subjects

A Kaplan-Meier figure will be provided as well, for subjects in the double-blind safety set, to display a comparison of time to first sedative TEAE in Study Part A between the treatment arms. Other AE outputs may be added, as appropriate, and documented in the clinical study report.

Summaries involving severity (or relationship to study drug) will use the most severe (or most related) event when a subject has more than one event for a term. If the severity (or relationship to study drug) of an AE is missing, then the AE will be considered as severe (or related).

For subjects in the open-label safety set, AEs with missing start dates (and no evidence of a stop date prior to first dose of open-label dosing) will be considered as treatment-emergent for the open-label study part only. Otherwise, if the AE has a stop date during the double-blind study part, then the AE will be considered as treatment-emergent only for the double-blind study part. For subjects in the double-blind safety set who are not in the open-label safety set, AEs with missing start dates (and no evidence of a stop date prior to first dose of double-blind dosing) will be considered as treatment-emergent for the double-blind study part only. In general, partial start and stop dates will be handled conservatively.

6.6.4 Other Safety Analysis

6.6.4.1 Clinical Laboratory Evaluations

Clinical safety laboratory evaluations include clinical hematology, serum chemistry, and urinalysis.

These evaluations will be summarized using SI units for baseline, post-dosing, and change from baseline for visits across both the double-blind and open-label study parts, using the double-blind and open-label safety sets, respectively.

In addition, shift tables showing the number of subjects with low, normal, or high values at each post-baseline visit according to the central laboratory's reference ranges will be presented. Only observations from analysis visits corresponding to scheduled visits of a given evaluation will be presented in these summaries.

Only the scheduled parameters will be included in the summaries.

In addition, the number and percentage of subjects with potentially clinically important (PCI) laboratory values will be presented for both the double-blind and open-label study parts, using the corresponding analysis sets. The percentages will be calculated relative to the number of subjects with available baseline values and at least 1 post-baseline assessment. These PCI summaries will consider repeat and unscheduled assessments, as well. Criteria for potentially clinically important values are presented below in Table 4, respectively. In addition, the number and percentage of subjects with potentially clinically important (PCI) in all three liver function assessments (i.e., ALT/ AST/ Bilirubin Total)) will be also summarized. Any changes in PCI criteria will be documented in the clinical study report (CSR), as applicable.

Table 4. Criteria for Potentially Clinically Important Values for Clinical Laboratory Parameters

Laboratory Category **Laboratory Parameter**	**PCI Criteria**
Hematology:	
Hemoglobin	< 10 g/dL or >17g/dL
Hematocrit	<32% or >47%
RBC	<2.5 x10^6/μL
WBC	<2.80 x10^3/μL or >13.50 x10^3μL
Neutrophils	<1.5 x10^3/μL or <39%
Lymphocytes	<10% or > 50%
Monocytes	>25%
Eosinophils	>10%
Basophils	>15%
Platelet Count	<100 or >600 (x10^3 /μL)
Clinical Chemistry:	
SGOT (AST)	> 2 x ULN
SGPT (ALT)	> 2 x ULN
Gamma Glutamyl Transpeptidase (GGT)	> 3 x ULN
Bilirubin	> 1.5 x ULN
LDH	>3 x ULN
Glucose, blood	<55 mg/dL or >160 mg/dL
Urea Nitrogen (BUN)	>30 mg/dL
Creatinine, serum	>1.5 x ULN
Sodium	<130 mEq/L (grade III) or >150 mEq/L
Potassium, serum/plasma	>6.0 mEq/L
Chloride	<90 or >115 mEq/L
Albumin	<2.8g/dL
Calcium	<8 mg/dL or >11.5 mg/dL
TSH	<0.50 or >6.4 μIU/mL
Total Protein, plasma or serum	<5.6 or >8.4 g/dL
Magnesium	<1.3 mEq/L
Phosphorus, inorganic	<2.0 mg/dL or >6.1 mg/dL

Cholesterol-H	>300 mg/dL
Uric acid, serum	>10 mg/dL Males and >8 mg/dL Females

Urinalysis:	
Protein	Positive Value (excluding trace)
Glucose	Positive Value (excluding trace)
Blood	Positive Value (excluding trace)
Ketones	Positive Value (excluding trace)
Bilirubin	Positive Value (excluding trace)

A by-subject listing of clinical laboratory data will be presented, for each visit across both study parts, with flags for potentially clinically important values, using the double-blind safety set.

6.6.4.2 Vital Signs

Descriptive statistics for baseline, post-baseline, and change from baseline will be computed by treatment arm for all vital sign parameters (weight, height, BMI, temperature, respiratory rate, supine pulse rate and blood pressure, and standing pulse rate and blood pressure) for visits across both the double-blind and open-label study parts, using the double-blind and open-label safety sets, respectively. Visits for vital signs evaluations will be presented as collected on the CRF (ie not derived via analysis visit windowing). These outputs will be repeated, stratified by age group (6 to 12, 13 to 17).

Postural orthostatic blood pressure values will be presented as well, which are defined at each visit as:

(first standing measurement) – (last supine measurement).

BMI values will be converted to percentiles using the Centers for Disease Control (CDC) Stature, Weight, and Body Mass Index-for-age gender specific charts, if available.

In addition, the number and percentage of subjects with potentially clinically important vital sign values will be presented for both the double-blind and open-label study parts, using the corresponding analysis sets. The percentages will be calculated relative to the number of subjects with baseline and at least 1 post-baseline assessment. These PCI summaries will consider repeat and unscheduled assessments, as well. Criteria for potentially clinically important values are presented below in Table 5. Any changes in PCI criteria will be documented in the CSR, as applicable.

Table 5. Criteria for Potentially Clinically Important Values for Vital Signs

Parameter	Unit	Lower Criteria	Upper Criteria
Pulse Rate	bpm	<=50	>=100
Supine Systolic Blood Pressure: 6-12 years	mm Hg	<90	>120
Supine Systolic Blood Pressure: 13-17 years	mmHg	<100	>140
Supine Diastolic Blood Pressure: 6-12 years	mm Hg	<50	>80
Supine Diastolic Blood Pressure: 13-17 years	mmHg	<60	>90
Standing Systolic Blood Pressure	mmHg		>130
Standing Diastolic Blood Pressure	mmHg		>95
Postural Orthostatic Hypertension-SBP (decrease)	mmHg		>=25
Postural Orthostatic Hypertension-DBP (decrease)	mmHg		>=15
Weight (change from baseline)	kg	7% decrease	7% increase
BMI Percentile	%	<5	>95

A figure will be presented, displaying BMI percentiles by treatment arm for visits in Study Part A and Study Part B separately, using the double-blind safety set and the open-label safety set, respectively. These figures will be stratified by Tanner Stage (genitals for males, breasts for females) at study part baseline (3 levels: 1, 2 to 4, and 5).

In addition, figures will be presented displaying observed mean change from baseline values by treatment arm, for pulse and blood pressure, for each scheduled post-baseline visit in Study Part A.

A by-subject listing of vital sign values will be presented, along with flags for potentially clinically important values, for each visit across both study parts, using the double-blind safety set.

6.6.4.3 12-Lead ECGs

Descriptive statistics for baseline, post-baseline, and change from baseline will be computed by treatment arm for all ECG parameters [HR, PR interval, QRS interval, QT interval, as measured; QTc intervals, as determined using the Fridericia correction (QTcF) and the Bazett correction (QTcB)] for visits across both the double-blind and open-label study parts, using the double-blind and open-label safety sets, respectively. As multiple ECGs were required at the baseline visit for each study part, the baseline value for that study part is defined as the mean of the readable ECG measurements obtained at that visit. For overall interpretation, it is the worst interpretation obtained at the baseline visit of the study part.

The number and percentage of subjects in each overall ECG interpretation category (normal, clinically significant abnormal, non-clinically significant abnormal) will be summarized for visits across both the double-blind and open-label study parts, using the corresponding analysis sets. Percentages will be based on the number of subjects with values at each visit. In addition,

shift tables showing the number of subjects with normal, clinically significant abnormal, or non-clinically significant abnormal ECG interpretation results at post-baseline visit will be provided. For the ECG interpretation summaries, if a subject has multiple interpretations at a particular visit, the most significant result will be selected for summary.

In addition, the number and percentage of subjects with potentially clinically important ECG values will be presented for both the double-blind and open-label study parts, using the corresponding analysis sets. The percentages will be calculated relative to the number of subjects with baseline and at least 1 post-baseline assessment. These PCI summaries will consider repeat and unscheduled assessments, as well. Criteria for potentially clinically important values are presented below in Table 6. Any changes in PCI criteria will be documented in the CSR, as applicable.

Table 6. Criteria for Potentially Clinically Important Values for ECG Results

Parameter	Unit	PCI Criteria
ECG Result		Abnormal (core lab) + CS (investigator)
Heart Rate	bpm	<=50 or >=100
PR Interval	msec	>=200
QRS Interval	msec	>=120
QT Interval	msec	>=480
QTcB Interval	msec	>=500
QTcF Interval	msec	>=500
QT Interval (Change from Baseline)	msec	>=30 and <60 or >=60
QTcB Interval (Change from Baseline)	msec	>=30 and <60 or >=60
QTcF Interval (Change from Baseline)	msec	>=30 and <60 or >=60
Rhythm		Any rhythm other than sinus rhythm (sinus includes Normal sinus rhythm, sinus arrhythmia, sinus tachycardia, sinus bradycardia) , which has overall interpretation as 'Abnormal'

A by-subject listing will be provided which shows ECG values and interpretations, along with flags for potentially clinically important values, for each visit across study parts, using the double-blind safety set.

6.6.5 Extent of Exposure and Compliance

For each study part, a subject's drug exposure (measured in days) will be defined as (date of last dose in study part – date of first dose in study part + 1). Drug exposure in weeks will be calculated by dividing the exposure in days by 7, and be summarized as a continuous variable. Subjects who do not take any study drug will have exposure of 0 days. The numbers and

percentages of subjects within exposure categories will also be presented by 6 week increments through 54 weeks (ie <1 week, 1 - <6 weeks, 6 - <12 weeks, etc.)

For each study part and period (ie Dose Optimization Period, Dose Maintenance Period, Dose Tapering Period), investigational product dosing compliance is defined as the total number of tablets (for TAK503) or capsules (for atomoxetine) actually taken by a subject during that period divided by the number of capsules expected to be taken during the same period multiplied by 100. The total number of tablets or capsules actually taken is calculated by the total number of tablets or capsules dispensed minus the number of tablets or capsules returned. If a bottle is not returned, the number of tablets or capsules returned for that bottle will be imputed to zero. The number of tablets or capsules expected to be taken is calculated as the number of days the subject was in the particular period multiplied by the number of tablets or capsules to be taken per day during that period. For subjects in the placebo arm, tablets plus capsules will be considered in compliance calculations.

Compliance rates will be summarized with descriptive statistics and subject counts within compliance categories (<80%, 80 to <120%, ≥120%) for the Dose Optimization Period, the Dose Maintenance Period, Dose Tapering Period, and the whole Study Part. Summaries of study drug exposure and compliance will be presented for subjects in the double-blind safety set for Study Part A (and its corresponding periods), as well as subjects in the open-label safety set for Study Part B (and its corresponding periods). These outputs will be repeated, stratified by site, age group (6 to 12, 13 to 17), race (White, Black or African American, Other) and sex.

A by-subject listing will present total drug accountability for each study part, using the respective safety set.

6.7 Pharmacokinetic, Pharmacodynamic, and Biomarker Analyses

Not applicable.

6.8 Interim Analyses

No interim analysis, adaptive design, or data monitoring committee is planned for this study.

7.0 REFERENCES

Drake, C., Nickel, C., Burduvali, E., Roth, T., Jefferson, C. and Pietro, B. 2003. The Pediatric Daytime Sleepiness Scale (PDSS): Sleep Habits and School Outcomes in Middle-school Children. *Sleep,* 26, 455-8.

DuPaul, G. J., Reid, R., Anastopoulos, A. D., Lambert, M., Watkins, M. and Power, T. 2016. Parent and teacher ratings of attention-deficit/hyperactivity disorder symptoms: Factor structure and normative data. *Psychol Assess,* 28, 214-25.

Guy, W. (ed.) 1976. *Clinical Global Impression (CGI),* Rockville, MD: US Department of Health, Education, and Welfare.

Conners 3rd Edition. Manual. Toronto, Ontario, Canada: Multi-Health Systems, Inc. 2008.

Hughes, C., Rintelmann, J., Emslie, G., Lopez, M. and MacCabe, N. 2001. A revised anchored version of the BPRS-C for childhood psychiatric disorders. *J Child Adolesc Psychopharmacol,* 11, 77-93.

Hughes, C. W. 2008. Guidelines for Administration and Use of the Brief Psychiatric Rating Scale for Children nine and 21-item versions. Dallas (Texas): UT Southwestern Medical Ctr.

Lingjaerde, O., Ahlfors, U. G., Bech, P., Dencker, S. J. and Elgen, K. 1987. The UKU side effect rating scale. A new comprehensive rating scale for psychotropic drugs and a cross-sectional study of side effects in neuroleptic-treated patients. *Acta Psychiatr Scand Suppl,* 334, 1-100.

Marshall, W. A. and Tanner, J. M. 1969. Variations in pattern of pubertal changes in girls. *Archives of disease in childhood,* 44, 291-303.

Marshall, W. A. and Tanner, J. M. 1970. Variations in the pattern of pubertal changes in boys. *Archives of disease in childhood,* 45, 13-23.

Overall, J. E. and Pfefferbaum, B. 1982. The Brief Psychiatric Rating Scale for Children. *Psychopharmacol Bull,* 18, 10-6.

Ouyang, J., Carrol, K.J., Koch, G., and Li, J. 2017. Coping with Missing Data in Phase III Pivotal Regression Trials: Tolvaptan in Subjects with Kidney Diseases, a Case Study. *Pharmaceutical Statistics*, 16 (4), 250-266

Riley AW, Robertson JA, Forrest CB, et al. *Technical Manual for the Child Health and Illness Profile–Child Edition (CHIP-CETM) Parent and Child Report Forms*, 1.0 ed. Baltimore: Johns Hopkins University; 2001.

Shang, C.-Y. and Gbau, S. S.-F. 2012. Improving Visual Memory, Attention, and School Function with Atomoxetine in Boys with Attention-Deficit/Hyperactivity Disorder. *Journal of Child and Adolescent Psychopharmacology,* 22, 353-363.

8.0 CHANGES TO PROTOCOL PLANNED ANALYSES

The APRS assessment was removed as a secondary efficacy endpoint with Protocol Amendment 5, due in large part to the COVID-19 pandemic, since many subjects enrolled prior to Amendment 5 are missing APRS data. APRS data that has been collected will be summarized descriptively (total and subscale scores) as well as presented in a listing. Since the APRS waiver

was implemented at a distinct timepoint, impacting all subjects equally and simultaneously (e.g., systemic bias), subsequent missing APRS assessments will be assumed to be missing at random. No imputation will be performed for missing APRS assessments. Unless data permit, no inferential analysis of this endpoint will be performed. Missing APRS assessments will not be a reason for exclusion from any analysis set.

Starting from Protocol Amendment 5, the C3PS total score, as well as analysis of Learning Problems and Executive Functioning subscales, has been added as a secondary efficacy endpoint as an alternative approach to obtain schoolwork performance related data following removal of APRS as a secondary efficacy endpoint. Limitations related with the use of this parents scale instead of scales for teachers more comprehensive and appropriate to assess school performance, will be discussed in the final results of this secondary efficacy endpoint.

As of Protocol Amendment 7, TAK-503 and atomoxetine will be evaluated after up to 12 months of double-blinded treatment using results from the CANTAB RTI task via the MMRM; however, inferential hypothesis testing via the noninferiority framework at the Week 49A timepoint will no longer be performed due to challenges with retaining an evaluable sample size that would allow for sufficient power for this comparison; superiority to placebo comparisons of TAK-503 and atomoxetine at the 18-week timepoint, separately, will still be performed, as in the original protocol.

9.0 APPENDIX

9.1 Changes From the Previous Version of the SAP

- Addition in 6.5.2.1 to include a figure for ADHD-RS-5 analysis
- Addition in 6.6.1.2 to include a figure for CANTAB RTI analysis
- Addition in 6.6.1.3 to include an additional sensitivity analysis
- Editorial changes to section 6.6.3 for AE analysis by weight-adjusted dose
- Addition in 6.6.4.1 to include a summary analysis for PCI of all three liver function assessments
- Addition in 6.6.4.2 to include figures for pulse and blood pressure.
- Further clarification for Rhythm in Table 6.

9.2 Data Handling Conventions

9.2.1 General Data Reporting Conventions

In general, summary statistics will be presented to 1 more decimal place than the recorded data. The standard deviations (SDs) will be presented to 2 more decimal places than the recorded data. Confidence intervals about a parameter estimate will be presented using the same number of decimal places as the parameter estimate. In general, the minimum and maximum values will be presented to the same number of decimal places as the recorded data. Values collected via the

CamCog electronic device may be presented to fewer decimals than collected (up to 3 decimal places), as appropriate.

P-values will be generally presented to 3 decimal places. P-values less than 0.001 will be reported as <0.001.

Data collected at unscheduled visits will be listed, and may be included in summary tables via analysis visit windowing, if applicable.

In the tables and listings, individual subjects will be identified by a combination of site number and screening number, e.g. XX-YYY where XX is the site number and YYY is the screening number. All collected data will be presented in data listings, and the listings will be sorted by treatment arm, subject, and time point where applicable.

9.2.2 Definition of Baseline

The "baseline" assessment for Study Part A will be defined as the last observed value before the first dose of blinded IMP. For subjects who enter Study Part B, the "baseline" assessment for analyses using the open-label safety set is the last observed value before the first dose of open-label TAK-503, including consideration of values from Study Part A, unless otherwise specified.

9.2.3 Definition of Visit Windows

For Study Part A, analysis day 1 is defined as the date on which a subject is administered their first dose of the blinded IMP. Other analysis days in Study Part A are defined relative to analysis day 1 with day -1 being the day prior to analysis day 1. Analysis days for Study Part B are defined in an analogous manner.

Unless otherwise specified, a windowing convention will be used to determine the analysis visit for a given study visit for observed data summaries and analyses of efficacy and safety endpoints. Table 7 below outlines the windowing convention for subjects in either of the active arms for analysis period A (including visits beyond Week 18A for which data were collected prior to those visits being eliminated under Protocol Amendment 7), while Table 8 outlines the windowing convention for subjects in the placebo arm for analysis period A. Table 9 outlines the windowing convention for all subjects for analysis period B.

Unless otherwise specified, if a subject has more than one measurement in the same visit window, the measurement closest to the target day will be used. If two measurements in the same window are of equal distance to the target day, the measurement that occurs after the target day will be used. If two or more measurements occur on the same day, the last repeat value will be used.

Table 7 Analysis Visit Windows for Endpoints in Analysis Period A- Active Arms

Visit ID (Analysis Visit)	Target Day (Analysis Day)	Window (Analysis Day)
Visit 3A/Week 1A	7	Days 1 to 11
Visit 4A/Week 2A	14	Days 12 to 18
Visit 5A/Week 3A	21	Days 19 to 25

Visit 6A/Week 4A	28	Days 26 to 32
Visit 7A/Week 5A	35	Days 33 to 39
Visit 8A/Week 6A	42	Days 40 to 46
Visit 9A/Week 7A	49	Days 47 to 60
Visit 10A/Week 10A	70	Days 61 to 98
Visit 11A/Week 18A	126	Days 99 to 144
Visit 12A/Week 23A	161	Days 145 to 207
Visit 13A/Week 36A	252	Days 208 to 298
Visit 14A/Week 49A	343	≥Day 299

Table 8 Analysis Visit Windows for Endpoints in Analysis Period A- Placebo Arm

Visit ID (Analysis Visit)	Target Day (Analysis Day)	Window (Analysis Day)
Visit 3A/Week 1A	7	Days 1 to 11
Visit 4A/Week 2A	14	Days 12 to 18
Visit 5A/Week 3A	21	Days 19 to 25
Visit 6A/Week 4A	28	Days 26 to 32
Visit 7AWeek 5A	35	Days 33 to 39
Visit 8A/Week 6A	42	Days 40 to 46
Visit 9A/Week 7A	49	Days 47 to 60
Visit 10A/Week 10A	70	Days 61 to 98
Visit 11A/Week 18A	126	Days 99 to 144

Table 9 Analysis Visit Windows for Endpoints in Analysis Period B

Visit ID (Analysis Visit)	Target Day (Analysis Day)	Window (Analysis Day)
Visit 3B/Week 1B	7	Days 1 to 11
Visit 4B/Week 2B	14	Days 12 to 18
Visit 5B/Week 3B	21	Days 19 to 25
Visit 6B/Week 4B	28	Days 26 to 32
Visit 7B/Week 5B	35	Days 33 to 39
Visit 8B/Week 6B	42	Days 40 to 46
Visit 9B/Week 7B	49	Days 47 to 60
Visit 10B/Week 10B	70	Days 61 to 147
Visit 11B/Week 23B	224	Days 148 to 238
Visit 12B/Week 36B	252	Days 239 to 298
Visit 13B/Week 49B	343	≥ Day 299

9.3 Analysis Software

All statistical analyses, unless otherwise specified, will be conducted using SAS® Version 9.3, or later.