

NCT02372799

Study ID: VLZ-MD-22

Title: A Multicenter, Double-blind, Placebo- and Active-Controlled Parallel-Group Evaluation of the Safety and Efficacy of Vilazodone in Pediatric Patients With Major Depressive Disorder

Statistical Analysis Plan Amendment 1 Date: 05 October 2018

1.0

TITLE PAGE



**Forest Research Institute, Inc., a Subsidiary of Actavis plc
Harborside Financial Center, Plaza V
Jersey City, NJ 07311**

VLZ-MD-22

**A Multicenter, Double-Blind, Placebo- and Active-Controlled Parallel-group
Evaluation of the Safety and Efficacy of Vilazodone in Pediatric Patients with Major
Depressive Disorder**

STATISTICAL ANALYSIS PLAN

**Final: 05 May 2015
Amendment #1: 5 October 2018**

Confidentiality Statement

This document is the property of Forest Research Institute, Inc., and may not—in full or part—be passed on, reproduced, published, distributed to any person, or submitted to any regulatory authority without the express written permission of Forest Research Institute, Inc.

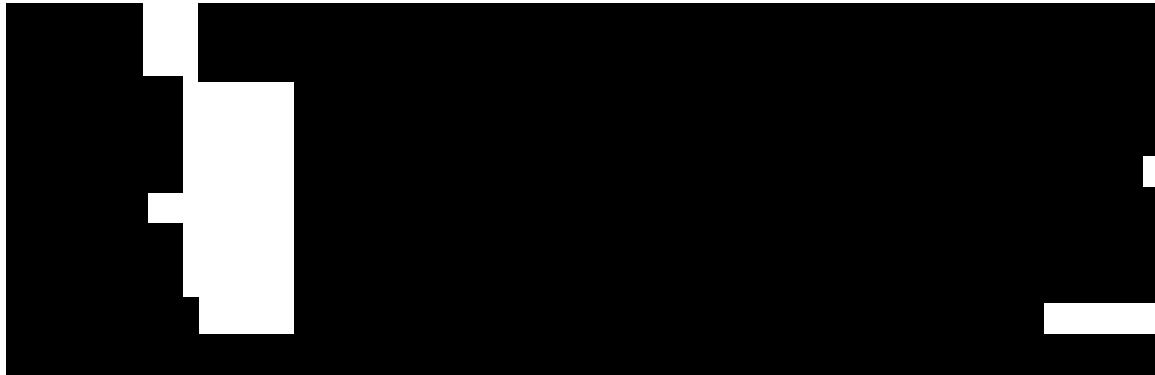
2.0 OVERALL TABLE OF CONTENTS

| | | |
|-------------|--|-----------|
| 1.0 | TITLE PAGE | 1 |
| 2.0 | overall table of contents | 2 |
| 2.1 | list of in-text tables | 4 |
| 2.2 | list of in-text figures..... | 5 |
| 3.0 | LIST OF ABBREVIATIONS | 6 |
| 4.0 | INTRODUCTION | 8 |
| 5.0 | OBJECTIVES | 13 |
| 6.0 | PATIENT POPULATIONS..... | 14 |
| 6.1 | SCREENED POPULATION..... | 14 |
| 6.2 | RANDOMIZED POPULATION..... | 14 |
| 6.3 | SAFETY POPULATION | 14 |
| 6.4 | INTENT-TO-TREAT POPULATION | 14 |
| 7.0 | PATIENT DISPOSITION | 15 |
| 8.0 | DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS | 16 |
| 9.0 | EXTENT OF EXPOSURE AND TREATMENT COMPLIANCE | 18 |
| 9.1 | EXTENT OF EXPOSURE | 18 |
| 10.0 | EFFICACY ANALYSES | 19 |
| 10.1 | PRIMARY EFFICACY PARAMETER..... | 20 |
| 10.2 | SECONDARY EFFICACY PARAMETER(S)..... | 21 |
| 11.0 | SAFETY ANALYSES..... | 22 |
| 11.1 | ADVERSE EVENTS..... | 22 |
| 11.2 | CLINICAL LABORATORY PARAMETERS | 24 |
| 11.3 | VITAL SIGNS..... | 27 |
| 11.4 | ELECTROCARDIOGRAM | 28 |
| 12.0 | HEALTH OUTCOMES ANALYSES | 31 |
| 13.0 | BLINDED INTERIM ANALYSIS..... | 32 |
| 14.0 | DETERMINATION OF SAMPLE SIZE | 33 |
| 15.0 | STATISTICAL SOFTWARE..... | 34 |

| | | |
|-------------|---|-----------|
| 16.0 | DATA HANDLING CONVENTIONS | 35 |
| 16.1 | VISIT TIME WINDOWS..... | 35 |
| 16.2 | DERIVED EFFICACY VARIABLES | 36 |
| 16.3 | AGE-AND-GENDER-CORRELATED VALUES FOR WEIGHT AND HEIGHT..... | 36 |
| 16.4 | REPEATED OR UNSCHEDULED ASSESSMENTS OF SAFETY PARAMETERS ... | 37 |
| 16.5 | MISSING DATE OF THE LAST DOSE OF INVESTIGATIONAL PRODUCT..... | 37 |
| 16.6 | MISSING SEVERITY ASSESSMENT FOR ADVERSE EVENTS | 38 |
| 16.7 | MISSING CAUSAL RELATIONSHIP TO INVESTIGATIONAL PRODUCT FOR ADVERSE EVENTS..... | 38 |
| 16.8 | MISSING DATE INFORMATION FOR ADVERSE EVENTS | 38 |
| | [REDACTED] | |
| 16.10 | CHARACTER VALUES OF CLINICAL LABORATORY PARAMETERS..... | 41 |
| 17.0 | CHANGES TO ANALYSES SPECIFIED IN PROTOCOL..... | 43 |
| 18.0 | REFERENCES | 44 |
| 19.0 | APPENDICES | 45 |
| | APPENDIX I PATTERN-MIXTURE MODEL DETAILS..... | 45 |

2.1

LIST OF IN-TEXT TABLES



2.2 LIST OF IN-TEXT FIGURES

| | |
|-------------------------------|----|
| Figure 4-1. Study Design..... | 10 |
|-------------------------------|----|

3.0 LIST OF ABBREVIATIONS

| | |
|-----------|--|
| AE | adverse event |
| ALP | alkaline phosphatase |
| ALT | alanine aminotransferase |
| ANCOVA | analysis of covariance |
| ANOVA | analysis of variance |
| AST | aspartate aminotransferase |
| BP | blood pressure |
| CDRS-R | Children's Depression Rating Scale-Revised |
| CGI-I | Clinical Global Impression - Improvement |
| CGI-S | Clinical Global Impression - Severity |
| CMH | Cochran-Mantel-Haenszel |
| CRF | case report form |
| C-SSRS | Columbia-Suicide Severity Rating Scale |
| DSM-IV-TR | Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision |
| ECG | electrocardiogram, electrocardiographic |
| eCRF | electronic case report form |
| ET | early termination |
| GLMM | generalized linear mixed model |
| IP | investigational product |
| ITT | intent to treat |
| LAR | legally authorized representative |
| LOCF | last observation carried forward |
| MDD | major depressive disorder |
| MedDRA | Medical Dictionary for Regulatory Activities |

| | |
|------|---|
| MMRM | mixed-effects model for repeated measures |
| OC | observed cases |
| PCS | potentially clinically significant |
| PID | patient identification |
| PK | pharmacokinetic |
| QTc | QT interval corrected for heart rate |
| QTcB | QT interval corrected for heart rate using the Bazett formula |
| QTcF | QT interval corrected for heart rate using the Fridericia formula |
| SAE | serious adverse event |
| SAP | statistical analysis plan |
| SI | Le Système International d'Unités (International System of Units) |
| TBL | total bilirubin |
| TEAE | treatment-emergent adverse event |
| ULN | upper limit of normal laboratory reference range |

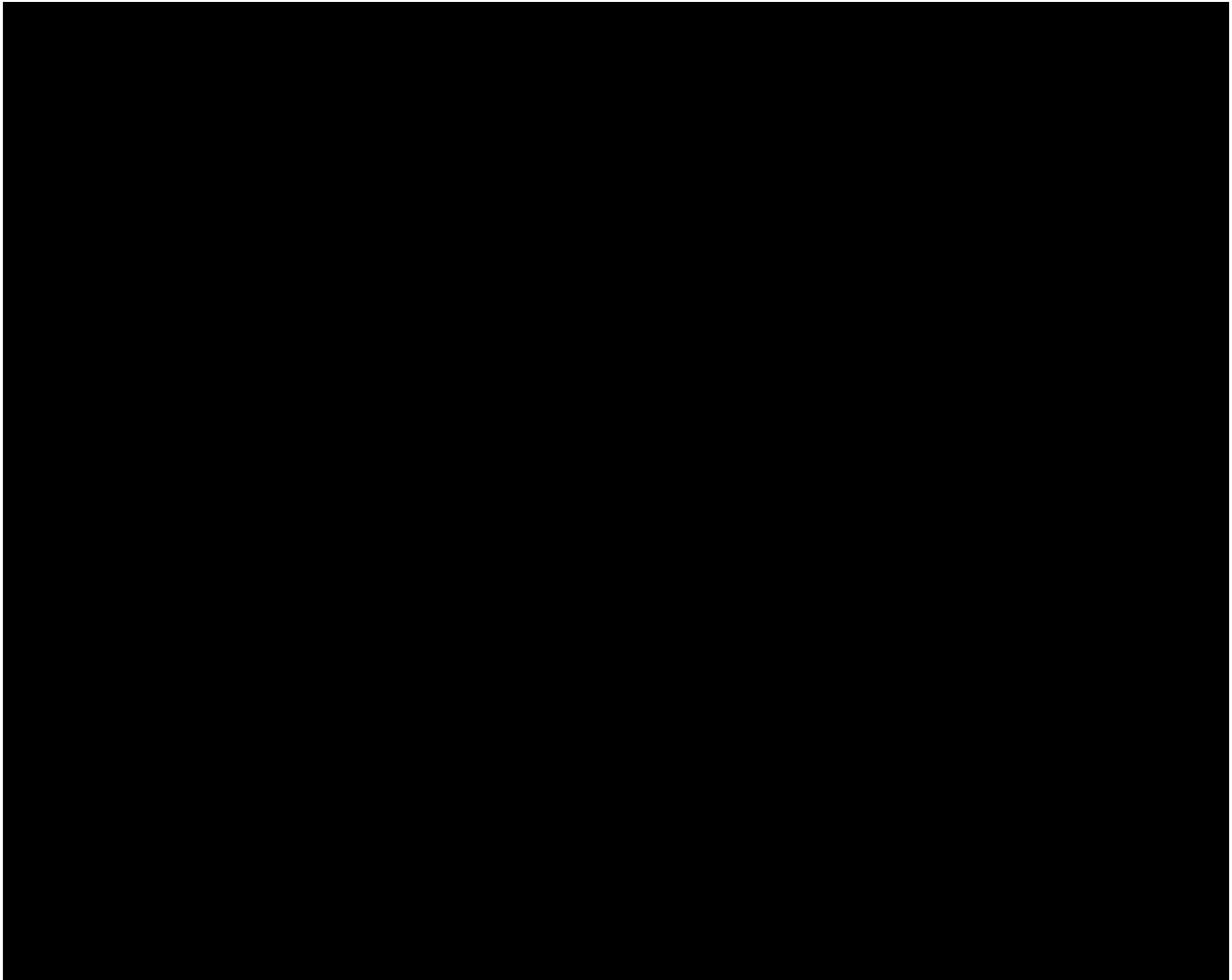
4.0 INTRODUCTION

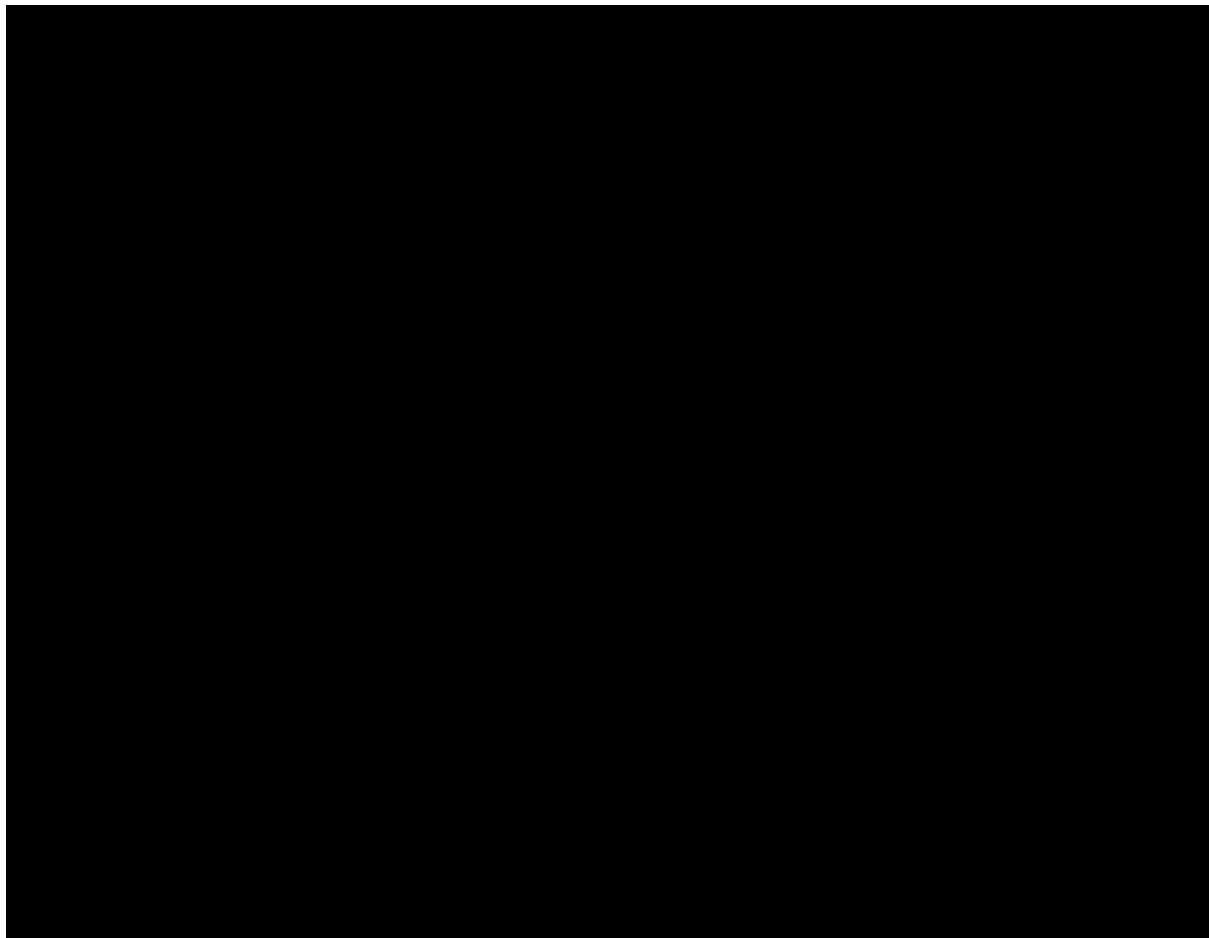
This statistical analysis plan (SAP) provides a more technical and detailed elaboration of the statistical analyses of the efficacy and safety data as outlined and/or specified in the protocol of Study VLZ-MD-22. Specifications of tables, figures, and data listings are contained in a separate document. The statistical analysis for pharmacokinetic/pharmacodynamic parameters will be specified in a separate document.

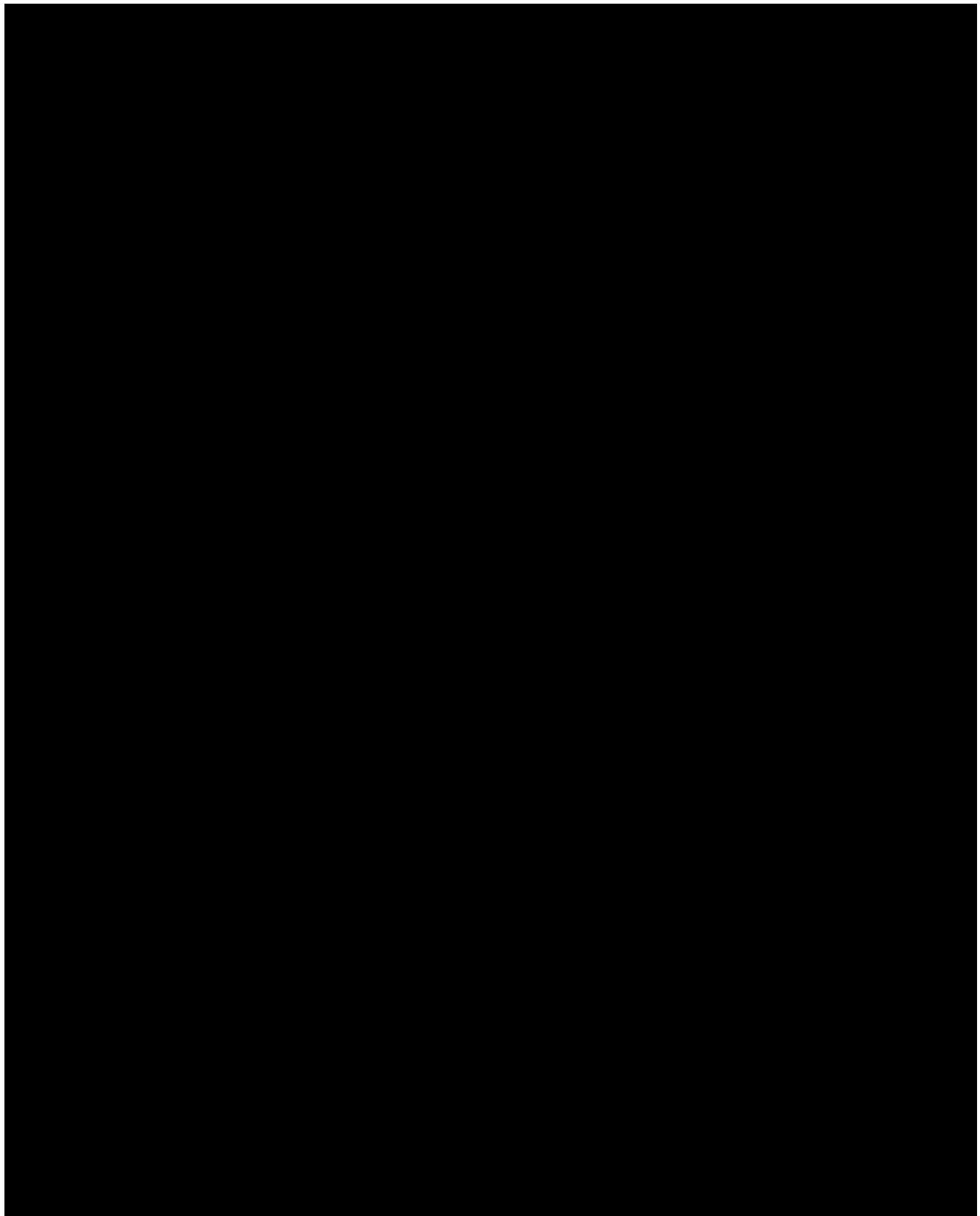
Study VLZ-MD-22 is a multicenter, randomized, double-blind, placebo- and active-controlled, parallel-group, flexible-dose study in pediatric patients, ages 7-17 years, who have been diagnosed with major depressive disorder (MDD) using the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR) diagnostic criteria for a minimum duration of 6 weeks. This study includes 3 treatment groups: placebo, vilazodone, and fluoxetine. A total of 400 patients (160 per group for placebo and vilazodone, 80 for fluoxetine) are planned to be randomized.

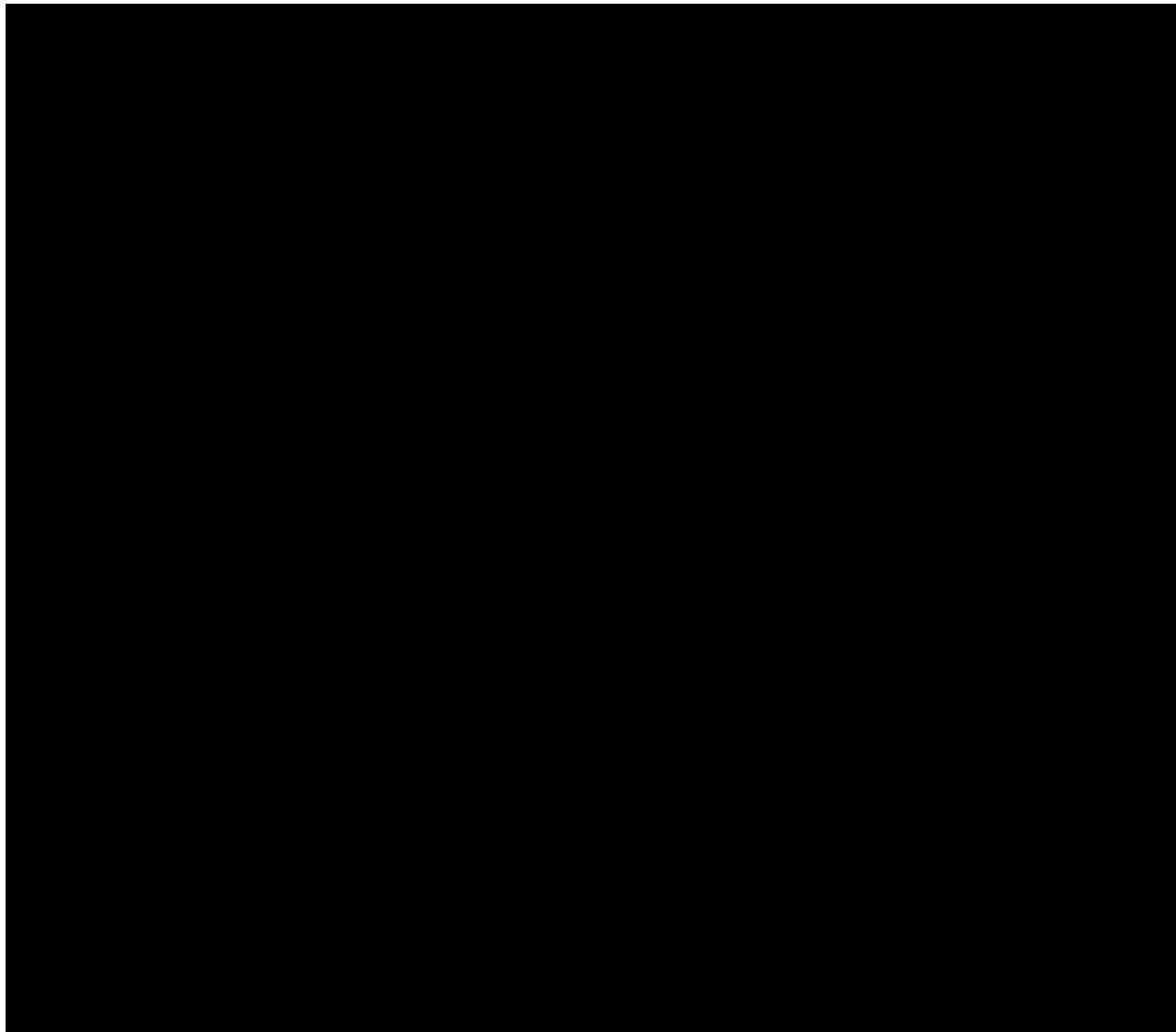
The study duration will be 10 weeks, starting with a 1-week screening period, followed by an 8-week of double-blind treatment period, and a 1-week of down-taper period. Patients must provide assent to participate and their parent/guardian/ legally authorized representative (LAR), and caregiver must provide written informed consent prior to the conduct of any study-specific procedures.











5.0 **OBJECTIVES**

The objective of this study is to evaluate the efficacy, safety, and tolerability of vilazodone compared with placebo in pediatric outpatients (7-17 years of age) with MDD.

In addition, the study is designed to obtain pharmacokinetic (PK) data to define the PK profile of vilazodone in the pediatric population (7-17 years of age).

6.0 PATIENT POPULATIONS

Four populations will be considered in the statistical analysis of the study.

6.1 SCREENED POPULATION

The Screened Population will consist of all patients who underwent a Screening Visit and received a screening number, and for whom the informed consent was obtained.

6.2 RANDOMIZED POPULATION

The Randomized Population will consist of all patients in the Screened Population who were randomized to a treatment group in the study.

6.3 SAFETY POPULATION

The Safety Population will consist of all patients in the Randomized Population who received at least 1 dose of double-blind investigational product.

6.4 INTENT-TO-TREAT POPULATION

The Intent-to-Treat (ITT) Population will consist of all patients in the Safety Population who had baseline and at least 1 postbaseline assessment of the Children's Depression Rating Scale-Revised (CDRS-R) total score.

7.0 PATIENT DISPOSITION

The number of patients in 3 of the study populations (Randomized, Safety, and ITT) will be summarized by treatment group and study center; the Screened Population will be summarized overall only by study center.

Screen-failure patients (i.e., patients screened but not randomized) and the associated reasons for failure to randomize will be tabulated overall for the Screened Population. The number and percentage of patients who complete the double-blind treatment period and of patients who prematurely discontinue during the same period will be presented for each treatment group and pooled across treatment groups for the Safety Population. The number and percentage of patients who enter the down-taper period and of patients who complete the down-taper period will be presented for each treatment group and pooled across treatment groups for the Safety Population.

The reasons for premature discontinuation from the double-blind treatment period as recorded on the disposition pages of the electronic case report forms will be summarized (number and percentage) by treatment group for the Safety Population. Percentage of premature discontinuations will be compared overall and for each discontinuation reason between treatment groups using Fisher's exact test. All patients who prematurely discontinue during the double-blind treatment period will be listed by discontinuation reason for the Safety Population.

8.0

DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS

Demographic parameters (age, sex, race, ethnicity) and other baseline characteristics (weight, height, body mass index) will be summarized descriptively by treatment group for the Safety and ITT populations, respectively. Baseline efficacy variables will be summarized by treatment group for the ITT Population.

Continuous variables will be summarized by number of patients and mean, SD, median, minimum, and maximum values. Categorical variables will be summarized by number and percentage of patients.

Comparability between treatment groups will be tested using a 2-way analysis-of-variance (ANOVA) model with treatment group and study center as the factors for continuous variables and a Cochran-Mantel-Haenszel (CMH) test, controlling for study center, for categorical variables.

Medical and surgical history/physical findings will be classified by system organ class and preferred term using the *Medical Dictionary for Regulatory Activities*, version 17.0 or newer. The number and percentage of patients with abnormalities in medical and surgical histories in each system organ class and preferred term will be summarized by treatment group for the Safety Population. Psychiatric history of major depressive disorder, other psychiatric history, prior treatment with psychotropic medication, and nondrug psychiatric treatment history will also be summarized by treatment group for the Safety Population.

The *World Health Organization Drug Dictionary*, version March 2014 or newer, will be used to classify prior and concomitant medications by therapeutic class and drug name.

Prior medication is defined as any recorded medication taken before the date of the first dose of double-blind investigational product. *Concomitant medication* is defined as any medication taken on or after the date of the first dose of double-blind investigational product.

Both prior and concomitant medications use will be coded by drug name and therapeutic class. The use of prior and concomitant medications will be summarized by the number and percentage of patients receiving each drug within each therapeutic class in each treatment group for the Safety Population. If a patient took a specific medication multiple times or took multiple medications within a specific therapeutic class, that patient would be counted only once for the coded drug name or therapeutic class.

Summaries for concomitant medication use will be presented for the double-blind treatment period and the down-taper period, separately. Any concomitant medications taken after the date of the last dose of double-blind investigational product in the study will not be presented in the summary tables but will be included in the patient data listings.

9.0 EXTENT OF EXPOSURE AND TREATMENT COMPLIANCE

9.1 EXTENT OF EXPOSURE

Exposure to double-blind investigational product for the Safety Population during the double-blind treatment period will be summarized in terms of treatment duration, calculated as the number of days from the date of the first dose of double-blind investigational product to the date of the last dose of double-blind investigational product during the double-blind treatment period, inclusive. Descriptive statistics (number of patients, mean, SD, median, minimum, and maximum) will be presented by treatment group. The number and percentage of patients will be presented for specific intervals of exposure by treatment group.

Patient-years, defined as total exposure to double-blind investigational product in years (excluding the down-taper period), will be summarized by treatment group for the Safety Population.

In addition, weekly and overall mean daily dose, overall modal daily dose, and final daily dose of investigational product during the double-blind treatment period will be summarized by treatment group for the Safety Population.

9.2 MEASUREMENT OF TREATMENT COMPLIANCE

[REDACTED]

10.0 **EFFICACY ANALYSES**

The efficacy analyses will be based on the ITT Population. *Baseline* for efficacy is defined as the last nonmissing efficacy assessment (at or prior to Visit 2) before the first dose of double-blind investigational product. All statistical tests will be 2-sided hypothesis tests performed at the 5% level of significance for main effects. All confidence intervals will be 2-sided 95% confidence intervals, unless stated otherwise.

For efficacy analyses in which study center is a factor, a small center will be defined as a center with fewer than 2 patients in any treatment group in the ITT Population. Small centers will be pooled to form pseudo-centers so that each treatment group includes at least 2 ITT patients within the center. Pooling will be done (within each country first) using the following algorithm:

Based on the number of ITT patients, small centers will be ordered from the largest to the smallest, and centers of the same size will be ordered from the largest center code to the smallest center code. The pooling process starts with the largest small center from the top, which will be pooled with the smallest from the bottom until a non-small pseudo center is formed. The process will be repeated using the small centers left out after the first pass. If any centers are left out at the end of the process, they will be pooled with the smallest pseudo center. If there is more than 1 smallest pseudo center, the pseudo center with the smallest center code will be selected. In case that the pseudo center formed by pooling all small centers is still a small center, it will be pooled with the smallest non-small center. If there is more than 1 smallest non-small center, the one with the smallest center code will be selected.

These pseudo-centers will be used for all efficacy analyses when the model is adjusted for study center.

By-visit analyses based on the mixed-effects model for repeated measures (MMRM) using the observed case (OC) approach will be performed for all continuous efficacy parameters with multiple post-baseline measurements.

In addition, by-visit analyses using the last-observation-carried-forward (LOCF) approach will be presented for all efficacy parameters. For the LOCF approach, only the postbaseline total score of a parameter will be imputed; individual item scores will not be carried forward to derive the total score. Baseline total score will be carried forward only for the intermittent missing scores immediately after baseline. If all the postbaseline values are missing, the baseline value will not be carried forward.

10.1 PRIMARY EFFICACY PARAMETER

The primary efficacy parameter is the change from baseline to Week 8 in CDRS-R total score. The primary analysis for comparing vilazodone vs placebo will be performed using an MMRM with treatment group, pooled study center, visit, and treatment-group-by-visit interaction as the fixed effects and the baseline value and baseline-value-by-visit interaction as the covariates. An unstructured covariance matrix will be used to model the covariance of within-patient scores. The Kenward-Roger approximation (Kenward and Roger, 1997) will be used to estimate denominator degrees of freedom. This analysis will only use the observed cases of postbaseline scores without imputation of missing values.

To control the overall type I error rate for the multiple comparisons across the primary and the secondary hypotheses, a fixed sequence procedure will be applied. Details are provided in Section [10.2](#).

In addition, two sensitivity analyses, LOCF and pattern-mixture model approaches, will be performed on the primary efficacy parameter. The LOCF approach is based on an analysis-of-covariance (ANCOVA) model including treatment group and pooled study center as factors and baseline CDRS-R total score as a covariate. Another sensitivity analysis using a pattern-mixture model approach based on non-future dependent missing value restrictions (Kenward et al., 2003) will be performed to assess the robustness of the primary MMRM results to the possible violation of the missing-at-random assumption. The details of this sensitivity analysis are as follows:

The pattern for the pattern-mixture model will be defined by the patient's last visit with an observed value. The observed CDRS-R total score at a visit is assumed to have a linear relationship with the patient's prior measurements. The missing values will be imputed under the assumption that the distribution of the missing observations differs from that of the observed only by a shift parameter value Δ . The dataset with missing values imputed will be analyzed using an ANCOVA model with treatment group and pooled study center as factors and baseline CDRS-R total score as a covariate for between-treatment group comparisons at Week 8. The imputation of missing values and the analysis will be performed multiple times and the inference of this sensitivity analysis will be based on the combined estimates using the standard multiple imputation technique. The range of values for the shift parameter Δ is selected as 0 to 6 based on experience with historical data.

More details of the proposed pattern-mixture model approach (e.g., the models for the pattern-specific identifiable densities and the unidentified conditional distributions, the shift parameter Δ , and the multiple imputation algorithm) are provided in Appendix I to this SAP.

To examine the consistency of the primary efficacy results across study centers, the treatment differences versus placebo treatment groups (mean \pm SE) in the change from baseline to Week 8 in CDRS-R total score will be graphed by pooled study center.

The comparison of fluoxetine vs. placebo will be performed to assess study assay sensitivity.

10.2 SECONDARY EFFICACY PARAMETER(S)

The secondary efficacy parameter is the change from baseline to Week 8 in CGI-S score, which will be analyzed using the MMRM approach similar to the one used for the primary efficacy parameter. A sensitivity analysis will also be performed using the LOCF approach as described in Section 10.1.

To control the overall family-wise type I error rate for multiple comparisons across the primary and the secondary efficacy parameters, the fixed sequence testing procedure will be implemented, i.e., the efficacy analysis of the secondary efficacy parameter will be carried out inferentially only if the null hypothesis for the primary efficacy parameter is rejected.

| Term | Percentage |
|------------|------------|
| GMOs | ~10% |
| Organic | ~75% |
| Natural | ~70% |
| Artificial | ~15% |
| Organic | ~75% |
| Natural | ~70% |
| Artificial | ~15% |
| Organic | ~75% |
| Natural | ~70% |
| Artificial | ~15% |
| Organic | ~75% |
| Natural | ~70% |
| Artificial | ~15% |
| Organic | ~75% |
| Natural | ~70% |
| Artificial | ~15% |

11.0 SAFETY ANALYSES

The safety analysis will be performed for the double-blind treatment period and the down-taper period (only for safety parameters collected during the down-taper period), separately, using the Safety Population unless stated otherwise.

The safety parameters include adverse events (AEs), clinical laboratory parameters, vital signs, electrocardiographic (ECG) parameters, suicide risk assessment using the Columbia-Suicide Severity Rating Scale (C-SSRS), and the change from baseline in the age and gender adjusted height (evaluation of growth). For each safety parameter, the last nonmissing safety assessment before the first dose of double-blind investigational product will be used as baseline for all analyses of that safety parameter. Continuous variables will be summarized by number of patients and mean, SD, median, minimum, and maximum values. Categorical variables will be summarized by number and percentage of patients.

For patients who take the down-taper investigational product, the double-blind treatment period starts with the first dose of double-blind investigational product and ends with the last assessment date up to the first dose of the down-taper investigational product. The double-blind down-taper period starts one day after the end of the double-blind treatment period and ends with the last dose of the down-taper investigational product or the last visit, whichever is later.

For patients who do not take the double-blind down-taper investigational product, the double-blind treatment period starts with the first dose of double-blind investigational product and ends with the last dose of the double-blind investigational product or the last visit, whichever is later.

11.1 ADVERSE EVENTS

Adverse events will be coded by system organ class and preferred term using the *Medical Dictionary for Regulatory Activities*, version 17.0 or newer.

An AE (classified by preferred term) that occurs during the double-blind treatment period or during the down-taper period will be considered a treatment-emergent adverse event (TEAE) if it was not present before the date of the first dose of double-blind investigational product, or was present before the date of the first dose of double-blind investigational product and increased in severity during the double-blind treatment period or during the down-taper period, respectively. If more than 1 AE was reported before the date of the first dose of double-blind investigational product and coded to the same preferred term, the AE with the greatest severity will be used as the benchmark for comparison with the AEs occurring during the double-blind treatment period or during the down-taper period that were also coded to that preferred term. An AE that occurs more than 30 days after the date of the last dose of double-blind investigational product in the study will not be counted as a TEAE but will be included in the listings.

An AE occurring during the down-taper period will be considered a newly emergent AE (NEAE) if it is not present before the start of the down-taper period or was present before the start of the down-taper period but increased in severity during the down-taper period. The NEAEs during the down-taper period will be summarized by body system, preferred term, and treatment group.

The number and percentage of patients reporting TEAEs in each treatment group will be tabulated by system organ class and preferred term and further categorized by severity and causal relationship to the investigational product, for both the double-blind treatment period and the double-blind down-taper period, separately. If more than 1 AE is coded to the same preferred term for the same patient, the patient will be counted only once for that preferred term using the greatest severity and strictest causality for the summarization by severity and by causal relationship to the investigational product, respectively.

The distribution of TEAEs by severity and causal relationship to the investigational product will be summarized by treatment group separately for the double-blind treatment period and the double-blind down-taper period, separately.

A common TEAE during the double-blind treatment period will be summarized in 2 ways: $\geq 1\%$ of patients in any treatment group and $\geq 2\%$ of patients in any treatment group. The incidence of common TEAEs will be summarized separately by preferred term and treatment group and will be sorted by decreasing frequency for the vilazodone treatment group.

A serious adverse event (SAE) that occurred on the date of the first dose of double-blind investigational product and up to 30 days after the date of the last dose of double-blind investigational product, inclusive, will be considered an on-therapy SAE. The number and percentage of patients who have on-therapy SAEs will be summarized by preferred term and treatment group and will be sorted by decreasing frequency for the vilazodone treatment group. In addition, the incidence of on-therapy SAEs that led to death will be summarized separately by treatment group and preferred term.

The number and percentage of patients in the Safety Population who have AEs leading to premature discontinuation of the investigational product will be summarized by preferred term and treatment group and will be sorted by decreasing frequency for the vilazodone treatment group.

Listings will be presented for patients with SAEs, patients with AEs leading to discontinuation, and patients who die (if any) for the Screened Population. All patients with SAEs, including those reported during the screening period or more than 30 days after the date of the last dose of the double-blind investigational product, and patients discontinuing because of AEs before taking double-blind investigational product will be included in these listings.

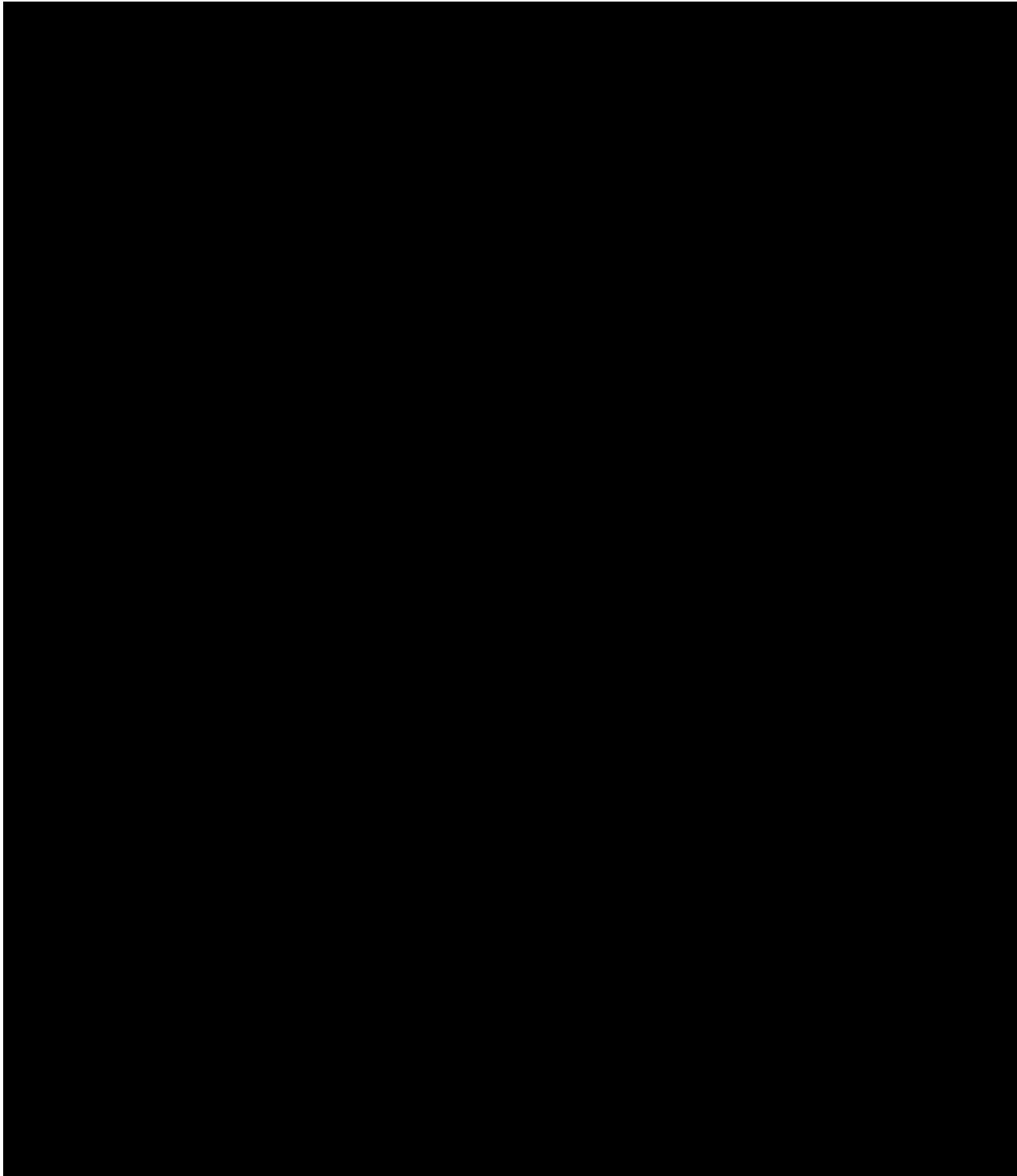
11.2 CLINICAL LABORATORY PARAMETERS

Descriptive statistics for clinical laboratory values (in SI units) and changes from the baseline values at each assessment time point (including the end of the double-blind treatment period) will be presented by treatment group for the following laboratory parameters:



[REDACTED]

[REDACTED]

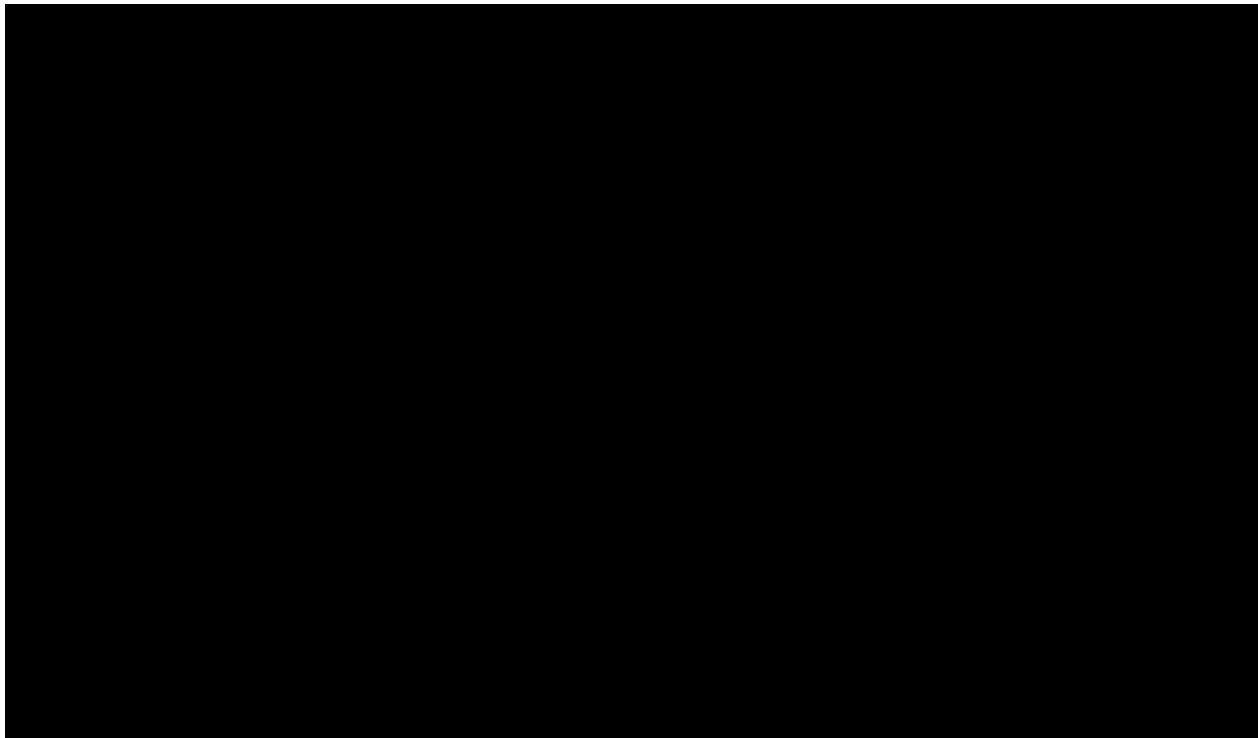


A horizontal bar chart consisting of 10 bars of varying lengths. Each bar is preceded by a small black square marker. The bars are arranged from top to bottom in the following order of lengths: very long, long, very long, long, medium, short, short, medium, long, very long.

11.3 VITAL SIGNS

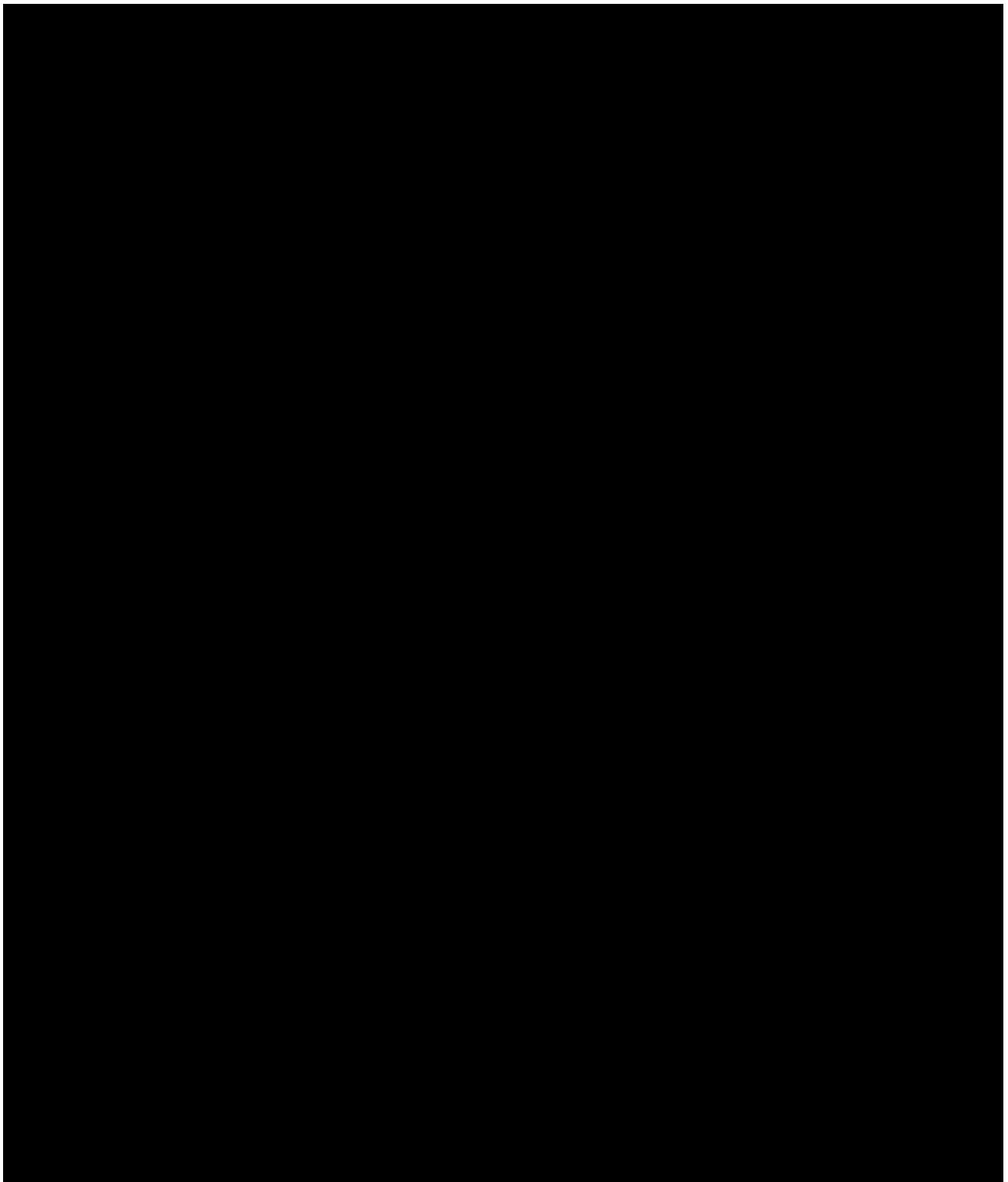
Descriptive statistics for vital signs (sitting radial pulse rate, sitting systolic and diastolic blood pressure, body weight and height) and their changes from baseline values at each visit, at the end of double-blind treatment period, and at the end of double-blind down-taper period will be presented by treatment group. Only patients with available baseline and at least 1 postbaseline assessment will be included in the summary.

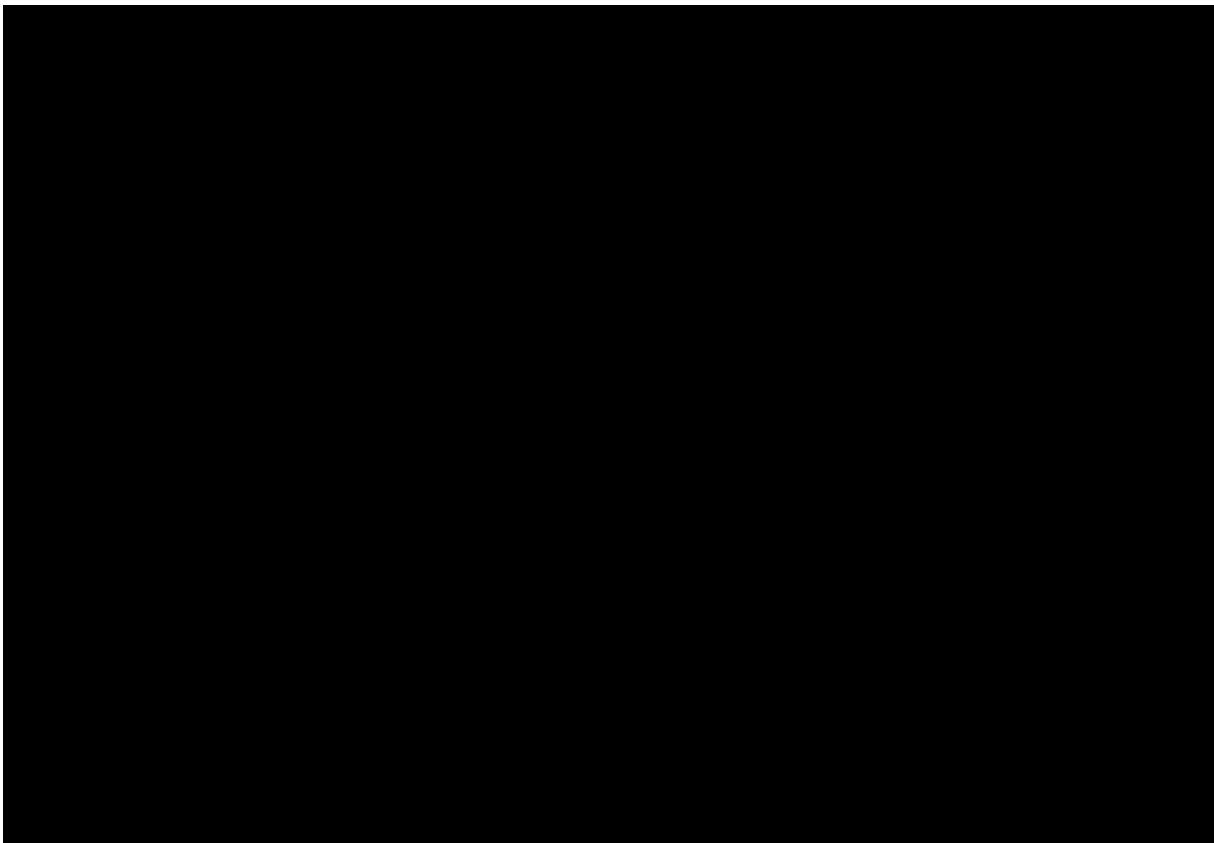
Vital sign values will be considered PCS if they meet both the observed-value criteria and the change-from-baseline criteria listed in [Table 11.3–1](#). The number and percentage of patients with PCS postbaseline values will be tabulated by treatment group separately for the double-blind treatment period and the double-blind down-taper period. The percentages will be calculated relative to the number of patients with available baseline and at least 1 postbaseline assessment. The numerator will be the total number of patients with available baseline values and at least 1 PCS postbaseline value. A supportive listing of patients with PCS postbaseline values will be provided, including the PID number, and baseline and all postbaseline (including non-PCS) values.



11.4 ELECTROCARDIOGRAM

Descriptive statistics for ECG parameters (heart rate, PR interval, QRS interval, RR interval, QT interval, and QTc interval) at baseline and changes from baseline values at each assessment time point (including the end of the double-blind treatment period) will be presented by treatment group. The QTc will be calculated using both the Bazett and Fridericia corrections. Only patients with available baseline and at least 1 postbaseline assessment will be included in the summary.





12.0 **HEALTH OUTCOMES ANALYSES**

Not applicable.

13.0 BLINDED INTERIM ANALYSIS

A blinded interim analysis will be conducted when approximately 75% of randomized patients have either completed the study or discontinued from the study. The blinded interim analysis is to obtain an estimate of the pooled standard deviation of the change from baseline to Week 8 in the CDRS-R total score. If the estimated pooled standard deviation is larger than the assumed pooled standard deviation specified in Determination of Sample Size (Section 14.0), the sample size may be increased to ensure an adequate power. However, due to the difficulties in recruiting pediatric patients with MDD, the total number of patients of the study will be capped at 500 (200 per treatment group for placebo and vilazodone, 100 for the fluoxetine treatment group). Detailed specification of the blinded interim analysis will be provided in the blinded interim analysis plan, a separate document. The interim analysis of sample size re-estimation will be performed by an independent statistician.

14.0 DETERMINATION OF SAMPLE SIZE

The primary efficacy parameter is the change from baseline to Week 8 during the double-blind treatment period in CDRS-R total score. A sample size of 400 patients (160 per group for placebo and vilazodone, and 80 for fluoxetine) was initially planned to provide 85% power to detect an effect size of 0.36 (treatment group difference of 4 units relative to pooled standard deviation of 11.1) for the primary efficacy parameter, based on an MMRM model using simulation method (Lu, 2012). The simulation assumed a correlation of 0.7 between the repeated measures, and a dropout rate of 17%, based on historical studies in pediatric patients with MDD.

Based on the planned blinded interim analysis which included the first 300 randomized patients (75% of the initial sample size), the study sample size was increased to a total of 455 patients (182 per group for placebo and vilazodone, and 91 fluoxetine patients) based on a pooled standard deviation of 12.3. During the teleconference with the FDA on the increased sample size, the FDA recommended a second interim analysis to be performed when the study enrolled the initially planned sample size of 400 patients. The second interim analysis included 427 randomized patients and the pooled standard deviation was 12.43. Therefore, a sample size of 470 patients (188 per group for placebo and vilazodone, and 94 for fluoxetine) is required to maintain the 85% statistical power to detect a treatment difference of 4 units for the primary efficacy endpoint.

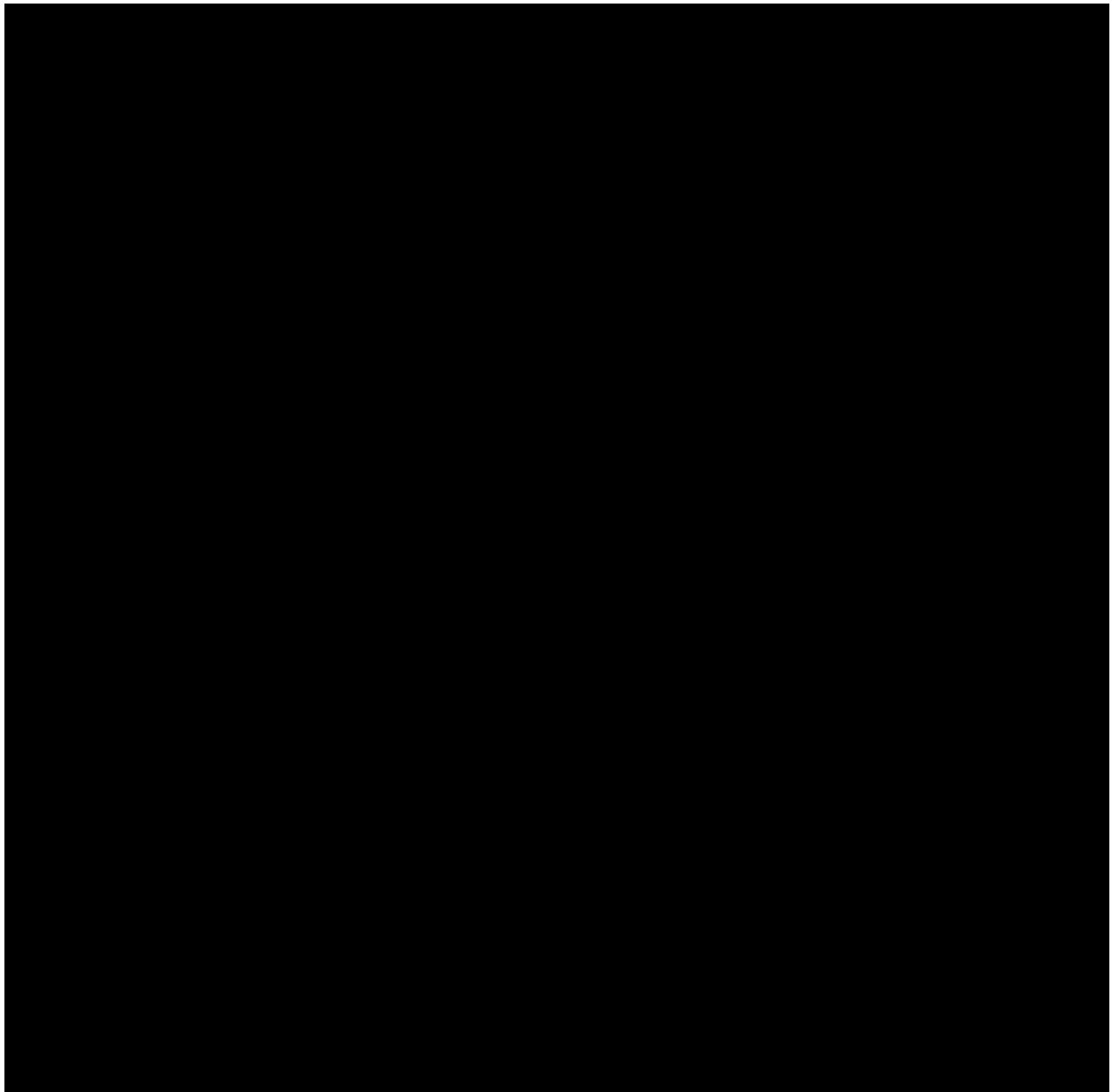
15.0 STATISTICAL SOFTWARE

Statistical analyses will be performed using version [REDACTED]
[REDACTED]

16.0 DATA HANDLING CONVENTIONS

16.1 VISIT TIME WINDOWS

[Table 16.1–1](#) and [Table 16.1–2](#) present the visits assigned for efficacy and safety analyses and the corresponding ranges of treatment days (window) during which an actual visit may occur.





16.2 DERIVED EFFICACY VARIABLES

The efficacy variables are derived as follows:

- *CDRS-R responders* are defined as patients with a $\geq 40\%$ reduction from baseline in CDRS-R total score. The value is 1 for CDRS-R responders and 0 otherwise
- *CDRS-R remitters* are defined as patients with $CDRS-R \leq 28$. The value is 1 for CDRS-R remitters and 0 otherwise

The total score at a particular visit will be calculated using $(\text{sum of nonmissing items}) \times (\text{total number of items}) / (\text{number of nonmissing items})$ only if the number of missing items is ≤ 2 for the CDRS-R score.

If a patient misses a postbaseline visit or if his/her postbaseline visit is outside of the visit time window, a record for the scheduled visit will be imputed using the last observed nonmissing value immediately before the missing value. If the missing value occurs at Week 1, the baseline value will be carried forward for Week 1, provided that at least 1 subsequent postbaseline assessment is available. For a composite scale such as CDRS-R total score, individual items of the rating scale will not be carried forward. Only total scores will be carried forward using the LOCF approach.

16.3 AGE-AND-GENDER-CORRELATED VALUES FOR WEIGHT AND HEIGHT

To adjust weight (kg) and height (cm) for sex and age, one needs to compare them to standard reference values for the same sex and age group, which are available in the United States Growth Charts and can be downloaded from:
http://www.cdc.gov/growthcharts/percentile_data_files.htm

The z-score is calculated as below

$$z = \frac{(X / M)^L - 1}{SL}, \text{ if } L \neq 0 \text{ and}$$

$$z = \frac{\ln(X / M)}{S}, \text{ if } L = 0,$$

where X is the physical measurement (e.g. weight and height) and L, M and S are the values from the appropriate table corresponding to the age in months (or length/stature) and sex (1 = male; 2 = female). X must be in metric measurements (kilograms or meters). This is called LMS method (Cole TJ, 1990), and parameters L, M, and S are the Box-Cox transformation power, median, and standard deviation, respectively, in the reference data, which again are provided in the reference data tables.

16.4 REPEATED OR UNSCHEDULED ASSESSMENTS OF SAFETY PARAMETERS

If a patient has repeated assessments before the date of first dose of double-blind investigational product, the results from the final nonmissing assessment made before the date of the first dose of double-blind investigational product will be used as baseline. If end-of-double-blind-treatment period assessments are repeated or if unscheduled visits occur, the last nonmissing postbaseline assessment during the double-blind treatment period will be used as the end-of-double-blind-treatment-period assessment for generating summary statistics. Likewise, if end of double-blind down-taper period assessments are repeated or if unscheduled visits occur, the last non-missing assessment during the double-blind down-taper period will be used as the end-of-double-blind-down-taper-period assessment for generating summary statistics. However, all postbaseline assessments will be used to determine PCS values for laboratory parameters, vital signs and ECG parameters, and to determine most severe suicidal ideation and most severe suicidal behavior from C-SSRS. All assessments will be presented in the data listings.

16.5 MISSING DATE OF THE LAST DOSE OF INVESTIGATIONAL PRODUCT

When the date of the last dose of investigational product in the study taken during the double-blind treatment period is missing for a patient in the Safety Population, all efforts should be made to obtain the date from the Investigator. If it is still missing after all efforts, then the last available dosing record date will be used as the last dose date.

16.6 MISSING SEVERITY ASSESSMENT FOR ADVERSE EVENTS

If severity is missing for an AE that started before the date of the first dose of double-blind investigational product, an intensity of mild will be assigned. If severity is missing for an AE that started on or after the date of the first dose of double-blind investigational product, an intensity of severe will be assigned. The imputed values for severity assessment will be used for the incidence summary; the values will be shown as missing in the data listings.

16.7 MISSING CAUSAL RELATIONSHIP TO INVESTIGATIONAL PRODUCT FOR ADVERSE EVENTS

If the causal relationship to the investigational product is missing for an AE that started on or after the date of the first dose of double-blind investigational product, a causality of yes will be assigned. The imputed values for causal relationship to double-blind treatment will be used for the incidence summary; the values will be shown as missing in the data listings.

16.8 MISSING DATE INFORMATION FOR ADVERSE EVENTS

The following imputation rules only apply to cases in which the start date for AEs is incomplete (i.e., partly missing).

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Final

05 Oct 2018

Final

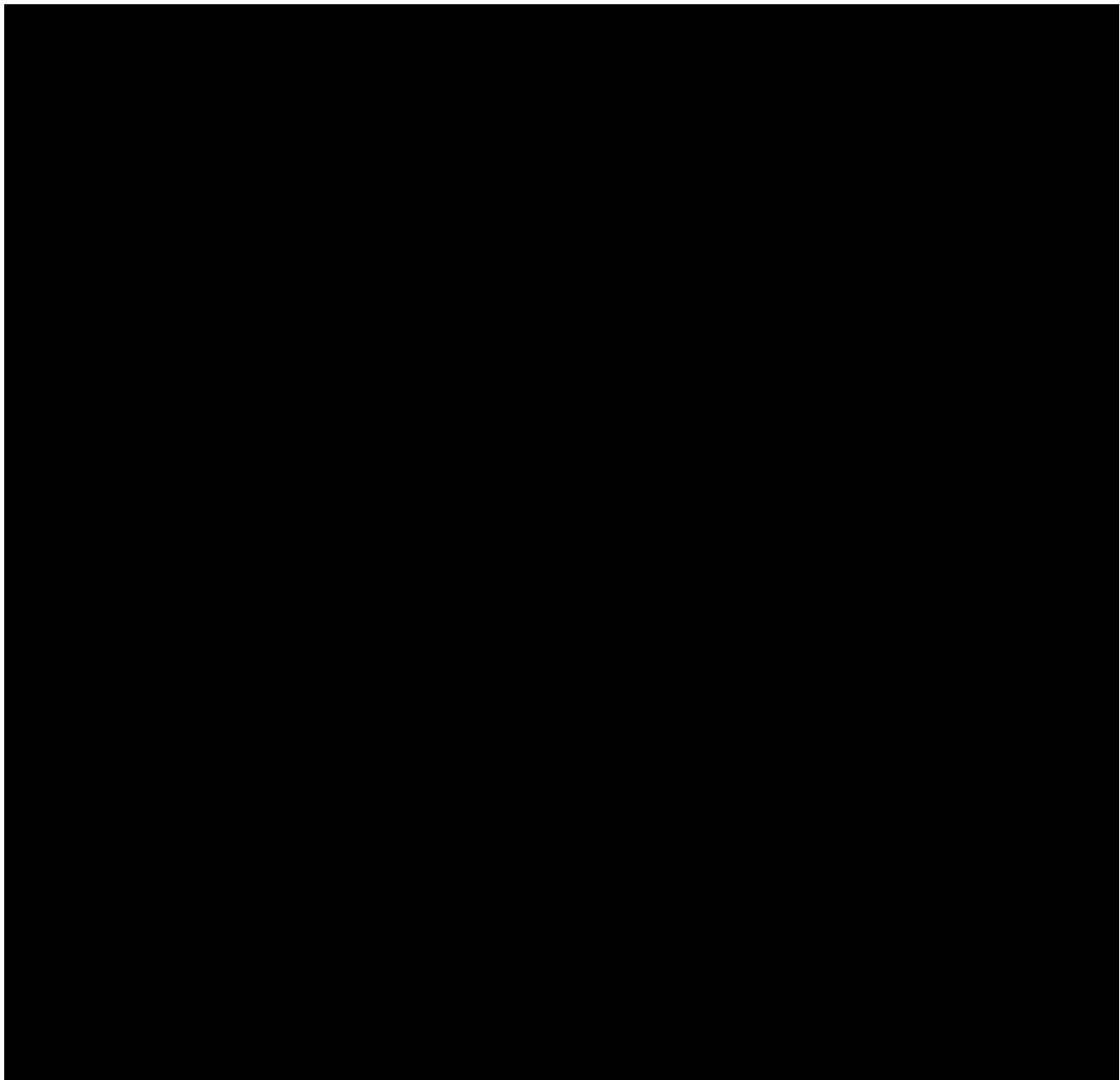
05 Oct 2018

The figure consists of seven bar charts arranged in a 7x2 grid. Each chart has 'Grade' on the x-axis (1, 2, 3, 4, 5, 6, 7) and 'Percentage' on the y-axis (0, 100). The charts show the percentage of students in each grade level who are in the 'High' category for different variables. The variables are: 'Number of hours spent on schoolwork per week', 'Number of hours spent on schoolwork per day', 'Number of hours spent on schoolwork per month', 'Number of hours spent on schoolwork per year', 'Number of hours spent on schoolwork per semester', 'Number of hours spent on schoolwork per term', and 'Number of hours spent on schoolwork per quarter'. The 'High' category is consistently the largest across all charts, peaking at approximately 80% for the first three charts and 70% for the last four charts.

| Grade | Number of hours spent on schoolwork per week | Number of hours spent on schoolwork per day | Number of hours spent on schoolwork per month | Number of hours spent on schoolwork per year | Number of hours spent on schoolwork per semester | Number of hours spent on schoolwork per term | Number of hours spent on schoolwork per quarter |
|-------|--|---|---|--|--|--|---|
| 1 | 80 | 75 | 70 | 70 | 70 | 70 | 70 |
| 2 | 80 | 75 | 70 | 70 | 70 | 70 | 70 |
| 3 | 80 | 75 | 70 | 70 | 70 | 70 | 70 |
| 4 | 75 | 70 | 70 | 70 | 70 | 70 | 70 |
| 5 | 75 | 70 | 70 | 70 | 70 | 70 | 70 |
| 6 | 75 | 70 | 70 | 70 | 70 | 70 | 70 |
| 7 | 75 | 70 | 70 | 70 | 70 | 70 | 70 |

16.10 CHARACTER VALUES OF CLINICAL LABORATORY PARAMETERS

If the reported value of a clinical laboratory parameter cannot be used in a statistical summary table because, for example, a character string is reported for a parameter of the numeric type, a coded value must be appropriately determined for use in the statistical analyses. The actual values, as reported in the database will be presented in the data listings.



17.0 CHANGES TO ANALYSES SPECIFIED IN PROTOCOL

No major changes have been made to the original protocol dated 26 Sep 2014.

18.0 **REFERENCES**

Cole TJ. The LMS method for constructing normalized growth standards. *Eur J Clin Nutr.*, 44: 45-60, 1990.

Kenward MG, Molenberghs G, Thijs H. Pattern-mixture models with proper time dependence. *Biometrika*, 2003;90:53-71.

Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 1997;53:983-997.

Lu K. Sample size calculations with multiplicity adjustment for longitudinal clinical trials with missing data. *Stat Med* 2012;31:19-28.

19.0

APPENDICES

APPENDIX I PATTERN-MIXTURE MODEL DETAILS

For repeated measures with the monotone missing mechanism, the pattern-mixture model with non-future dependent missing assumption, proposed by Kenward et al. (2003), provides a feasible solution to accommodate certain missing not at random (MNAR) mechanism. The methodology relies on constructing unidentifiable conditional densities using identifiable densities and borrows techniques from standard multiple imputation.

1. Non-Future Dependent Missing Assumption

Assume there are T designed visits in a longitudinal study and let y_i ($i = 1, 2, \dots, T$) represent patient's measurement at Visit i . When the missing mechanism is monotone, the pattern of missing data can be defined by the number of measurements (L) actually observed from the patient. Let $f(y_1, \dots, y_j | L = t)$ denote the conditional density of y_1, \dots, y_j , given that the last observed measurement is at Visit t . Then the overall density function for Pattern t can be written as

$$f(y_1, \dots, y_T | L = t) = f(y_1, \dots, y_t | L = t) f(y_{t+1} | y_1, \dots, y_t, L = t) \times \prod_{s=t+2}^T f(y_s | y_1, \dots, y_{s-1}, L = t) \quad (1)$$

Note on the right hand side of (1) the first factor is clearly identifiable from the observed data, while the second and the beyond are not, due to lack of available data. The second factor $f(y_{t+1} | y_1, \dots, y_t, L = t)$ could be identifiable based on an assumed relationship between $f(y_{t+1} | y_1, \dots, y_t, L = t)$ and $f(y_{t+1} | y_1, \dots, y_t, L \geq t + 1)$. The third and beyond factors $f(y_s | y_1, \dots, y_{s-1}, L = t)$ (with all $s \geq t + 2$) could be identifiable with the help of non-future dependent missing assumption.

For longitudinal data with dropouts, non-future dependent missing (NFD) mechanism (Kenward et al., 2003) assumes that the unidentifiable conditional distributions of y_s ($s \geq t + 2$), given earlier measurements, in Pattern t , is equal to the corresponding distribution in patterns $L \geq s - 1$:

$$f(y_s | y_1, \dots, y_{s-1}, L = t) = f(y_s | y_1, \dots, y_{s-1}, L \geq s - 1) \quad (2)$$

The right hand side of (2) can further be partitioned into

$$f(y_s | y_1, \dots, y_{s-1}, L \geq s-1) = \sum_{j=s-1}^T \omega_{s-1,j} f(y_s | y_1, \dots, y_{s-1}, L = j) \quad (3)$$

Where mixture probabilities $\omega_{s-1,j}$ are:

$$\omega_{s-1,j} = \frac{\alpha_j f(y_1, \dots, y_{s-1} | L = j)}{\sum_{t=s-1}^T \alpha_t f(y_1, \dots, y_{s-1} | L = t)}, \text{ and } \alpha_j \text{ represents the fraction of } \\ \text{patients from Pattern } j. \quad (4)$$

Each factor of the unidentifiable conditional distribution of y_s ($s \geq t+2$) on the right side of (1) can be expressed using the following:

- $f(y_s | y_1, \dots, y_{s-1}, L = s-1)$, the unidentifiable conditional distribution of the first missing in pattern $s-1$,
- $f(y_s | y_1, \dots, y_{s-1}, L = j)$, the identifiable conditional distributions of y_s given y_1, \dots, y_{s-1} of pattern j ($j \geq s$), and
- α_j , the fraction of patients from pattern j ($j \geq s-1$).

So under NFD, all the unidentifiable conditional distribution on the right side of (1) can be estimated and missing value could be therefore imputed based on the assumption for unidentifiable conditional distribution of the first missing.

We re-formulate the partition in (3), for $s \geq t+2$, as the following:

$$f(y_s | y_1, \dots, y_{s-1}, L = t) = \delta_{s-1} f(y_s | y_1, \dots, y_{s-1}, L = s-1) + (1 - \delta_{s-1}) f(y_s | y_1, \dots, y_{s-1}, L \geq s) \quad (5)$$

for $s \geq t+2$ with $\delta_{s-1} = \omega_{s-1,s-1}$.

Therefore, under monotone missing and NFD assumption, the unidentifiable conditional densities for Visit s in Pattern t ($s \geq t + 2$) can be expressed as a mixture distribution of $f(y_s | y_1, \dots, y_{s-1}, L = s - 1)$ - the unidentifiable conditional distribution of the first missing measurement y_s in Pattern $s - 1$, and $f(y_s | y_1, \dots, y_{s-1}, L \geq s)$ - the identifiable conditional distribution of y_s from all the patterns with observed data at Visit s or beyond:

$$f(y_s | y_1, \dots, y_{s-1}, L \geq s) = \sum_{j=s}^T \lambda_{s-1,j} f(y_s | y_1, \dots, y_{s-1}, L = j) \quad (6)$$

where the mixture probability

$$\lambda_{s-1,j} = \omega_{s-1,j} / (1 - \omega_{s-1,s-1}) = \frac{\alpha_j f(y_1, \dots, y_{s-1} | L = j)}{\sum_{t=s}^T \alpha_t f(y_1, \dots, y_{s-1} | L = t)} \text{ for } j \geq s, \text{ where } \alpha_j \text{ is the fraction of} \quad (7)$$

patients from Pattern j .

The conditional densities for the first missing are selected as:

$$f(y_s | y_1, \dots, y_{s-1}, L = s - 1) = f(y_s - \Delta | y_1, \dots, y_{s-1}, L \geq s) \text{ for } s = 2, \dots, T, \quad (8)$$

Note that the two distributions only differ by a shift (Δ) parameter. When $\Delta = 0$, the missing value y_s in Pattern $s - 1$ is imputed based on the distribution of all observed data up to Visit s , as a result, leading to missing at random (MAR) missingness. When $\Delta \neq 0$, (8) will introduce a scenario of MNAR. The similar idea was also presented in the recent publication “The Prevention and Treatment of Missing Data in Clinical Trials” by the National Academies Press. The selection of the plausible values for the shift parameter (Δ) is discussed in Section 3 of this appendix.

Note that per recommendation in Wang and Daniels (2011), only the observed data within pattern is assumed to be multivariate normal. The observed data distribution can be expressed in terms of the marginal distribution of baseline measurement and the conditional distributions of postbaseline measurements given earlier measurements. Assuming that these distributions are normal, the linear regression of each observation on prior observations will yield least-squares estimates of model parameters that can be utilized for independent posterior draws of model parameters for observed data. Multiple imputation approach will be used to estimate the overall mean at the final time point.

2. Imputation Procedure

All the missing data will be imputed to create complete datasets, then statistical analysis can be performed using appropriate techniques such as ANCOVA. The imputation can accommodate MNAR missing data mechanisms, based on the theory discussed in the previous section.

The model parameters for each dropout pattern, i.e., the mean, variance and proportions of observations in each pattern, are drawn from their posterior distributions prior to the imputation of missing data for a single imputation.

The details of imputation within a pattern, say Pattern t , are as the following:

Step 1. Impute the first missing value y_{t+1} for each patient in Pattern t ($t = 1, \dots, T - 1$):

- a. Compute estimates of mixture probabilities $\lambda_{s-1,j}$ in (7) with $s = t+1$ given the posterior draw of proportions of observations in each pattern and the posterior draw of regression parameters for the observed data.
- b. Draw a random integer from $\{s, \dots, T\}$ to index a component distribution on the right hand side of (6), using mixture probabilities obtained in a). Draw y_{t+1}^* from the identified component normal distribution. Impute the missing y_{t+1} as
$$\tilde{y}_{t+1} = y_{t+1}^* + \Delta.$$

Step 2. Impute the rest of the missing values of $y_{t+2}, y_{t+3}, \dots, y_T$ for patients in Pattern t :

Starting with imputation for y_{t+2} , first, similar to Step 1, draw y_{t+2}^* from the normal mixture (6) based on the observed y_1, \dots, y_t and the already imputed \tilde{y}_{t+1} for the patient.

Then the missing y_{t+2} is imputed as $\tilde{y}_{t+2} = y_{t+2}^* + \Delta$ with probability δ_{t+1} and as $\tilde{y}_{t+2} = y_{t+2}^*$ with probability $1 - \delta_{t+1}$, where the mixture probability $\delta_{t+1} = \omega_{t+1,t+1}$ is obtained from (4) given the posterior draw of proportions of observations in each pattern and the posterior draw of regression parameters for the observed data.

Missing values of y_{t+3} through y_T can be imputed similarly as y_{t+2} .

To summarize, the imputations of y_{t+1} through y_T is done recursively within each Pattern t (for all $t = 1, \dots, T - 1$) to create a complete dataset after imputation is done for all patterns with missing values.

The above imputation procedure is applied to all subjects in each missing data pattern to create a single imputed data set. Repeating the process of drawing parameters from the posterior distribution and imputing missing data given the posterior draw m times will yield m imputed data sets. The observed or imputed values at the final data point are averaged to obtain the overall mean estimate for each imputed data set, and the multiple imputation estimate is obtained by averaging the estimates across m imputations.

In this sensitivity analysis, m is set to equal to 20. The value of m is discussed in the context of imputation efficiency in standard multiple imputation theory (Rubin, 1987, p. 114), and $m = 20$ would provide at least 96% of relative efficiency as compared with using a large number of imputations (SAS/STAT User's Guide, p. 3796).

3. Determination of the Shift Parameter Values

The common shift parameter Δ is the difference between the mean of y_{t+1} among those who drop out at Visit t and those who remain beyond Visit t ($1 \leq t \leq T-1$). The exact value of Δ is unknown and can't be estimated from data because of missingness. The magnitude of Δ depends on the medical aspects of the trial. Using relevant historical data, one may select Δ as a proportion of the sample standard deviation or a proportion of observed treatment efficacy.

References:

Kenward MG, Molenberghs G, Thijs H. Pattern-mixture models with proper time dependence. *Biometrika*, 2003;90:53-71.

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

SAS Institute Inc. SAS/STAT 9.2 User's Guide, Second Edition. Cary, NC: SAS Institute Inc; 2009.

Wang C, Daniels MJ. A note on MAR, identifying restrictions, model comparison, and sensitivity analysis in pattern mixture models with and without covariates for incomplete data. *Biometrics*, 2011; 67(3):810-818.