Protocol A1281198

A PHASE 3, MULTICENTER, FOUR-WEEK, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED, PARALLEL-GROUP EFFICACY AND SAFETY TRIAL OF FLEXIBLE DOSES OF ORAL ZIPRASIDONE IN CHILDREN AND ADOLESCENTS WITH BIPOLAR I DISORDER (CURRENT OR MOST RECENT EPISODE MANIC)

Statistical Analysis Plan
(SAP)

Version: 1.2 (Amendment)

Author: PPD

Date: 21-JUL-2020

TABLE OF CONTENTS

LIST OF TABLES

APPENDICES

1. AMENDMENTS FROM PREVIOUS VERSION(S)

Amendment version 1.1 Changes:

- Update of sample size in Section 2.1 to include 14 additional subjects.
- Added a new sample size section (2.2) to explain the rater training issue and to update the sample size to address this issue and to update the power.
- Inclusion of sensitivity analysis, excluding the 14 subjects impacted by the e-rating in Section 8.2.2.
- Inclusion of subgroup analysis for subjects enrolled in Chinese sites.
- Updated rules for Treatment Misallocations in Section 5.4.
- Updated Appendix sections on SAS Code for the MMRM Analyses and Visit Window.

Amendment 1.2 Changes:

- Added a new sample size section (2.2.1) re-estimating the initial study design assumption for dropout rate; the above re-estimation, led to an update on section 8.2.2 calculations regarding the sensitivity analysis for primary endpoint.
- Update of sample size in Section 2.2.1 to include 18 subjects impacted by a rating training issue, instead of the previous 14 identified subjects. During evaluability meeting, four additional subjects were identified as impacted by a rater training issue affecting either the primary endpoint (YMRS) or the diagnostic interview (KSADS).
- Deleted entire section 8.2.4 of subgroup analysis for subjects enrolled in Chinese sites. Study did not enroll any China sites; this change led to an update to Table 2 to remove such tables for the summary of efficacy analyses.
- Updated Appendix sections 1.1 and CC on SAS code for the pattern analysis models CCI [redacted]
- Update section 8.2.7 for the Child Depression Rating Scale - Revised (CDRS-R) monitoring. A new listing will be provided for subjects with a CDRS-R raw score >=40 flagging any subjects with potential development of depression.
- Update section 8.2.5, removed M9 and M10 concentrations from analyses to be consistent with protocol amendment 2 dated 19 February2019.

2. INTRODUCTION

Ziprasidone is an atypical antipsychotic with a high affinity for the dopamine D_2 receptor and the $5HT2_A$ receptors. It blocks re-uptake of serotonin and norepinephrine and exhibits $5HT1_A$ agonist activity.

The present study is designed to assess the efficacy, safety, and tolerability of oral ziprasidone in child and adolescent subjects with Bipolar I Disorder (current or most recent episode manic). It is designed to comply with a Pediatric Research Equity Act (PREA) commitment to assess the safety and effectiveness of Geodon® (ziprasidone) as a treatment for bipolar disorder in pediatric patients ages 10 to 17 (children and adolescents), issued by the US Food and Drug Administration (FDA) to Pfizer in August 2004.

The overall goal of the study is to assess the safety profile and efficacy of ziprasidone in the treatment of pediatric mania in association with bipolar disorder, and to develop other relevant information, eg, CCI pertinent to evaluating the effects of ziprasidone in children and adolescents. The study population in this protocol will be limited to children and adolescents aged 10 to 17 years (inclusive). In this age range, the same DSM-V criteria for mania are employed for both the pediatric and adult population. While the lower end of the age range for bipolar disorder is not clear, it is both uncommon and difficult to make this diagnosis below the age of 10 years. On the other hand, bipolar disorder in the 10 to 17 year old population is thought to be relatively common and phenomenologically similar to bipolar disorder seen in adults.

2.1. Study Design

This will be a Phase 3, multicenter, 4-week, randomized, double-blind, placebo-controlled, parallel-group trial to evaluate the efficacy, safety, and tolerability of flexibly dosed ziprasidone compared with placebo for the treatment of Bipolar I Disorder (current or most recent episode manic) in children and adolescents aged 10 to 17 years (inclusive).

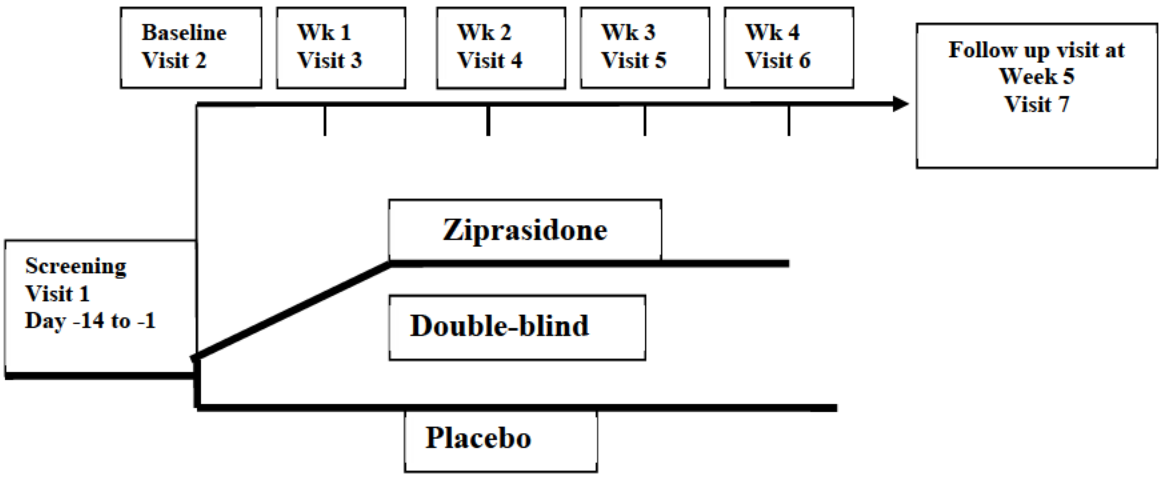
Ziprasidone (or placebo) will be administered as oral capsules given twice daily (BID) with food. Ziprasidone (or placebo) will be titrated over the first 7-14 days of treatment to the target dose range (60-80 mg/day for subjects <45 kg and 120-160 mg/day for subjects ≥45 kg), after which stable dosing will be maintained, if possible, through the end of the double-blind treatment period (end of Week 4, Day 29). However, depending on the clinical judgment of the investigator, the dose may be down titrated at any time in the case of tolerability or safety concerns.

Approximately 194 subjects (97 ziprasidone: 97 placebo) will be recruited from approximately 60 -70 worldwide sites. It is estimated that approximately 300 subjects will be required to be screened.

The study will begin with a screening visit to determine subject eligibility, followed by a 1-14 day period to allow for wash-out of exclusionary medications and the performance of two sequential pregnancy tests separated by a minimum interval of 14 days in all female subjects. Subjects who qualify will be randomized at baseline (Day 1) to receive either double-blind oral ziprasidone or placebo (randomization ratio 1:1). The randomization will be stratified by weight group (weight <45 kg, weight ≥45 kg).

Eligible subjects will remain in the treatment period for 4 weeks. After completion of the treatment phase, subjects will return for a post-treatment follow-up visit at week 5.

Subjects who are judged to have insufficient treatment response 1 week after completing their dose titration and have reached their maximum tolerated dose should be discontinued from the study and may be eligible to enroll in the open-label extension trial, provided there are no safety concerns and the minimum allowed dose per protocol is achieved.



Study medication will be dispensed at the baseline visit after all baseline assessments have been completed. A single titration card will be dispensed based on the subject's weight (≥45 kg or <45 kg) for the first week of dosing. The titration cards that will be used for the first week of the trial will have a maximum of 80 mg/day for subjects ≥45 kg and a maximum of 60 mg/day for subjects <45 kg. Subsequent dosing will follow as per the parameters of Section of the protocol. Dosing flexibility is allowed based on the clinical judgment of the investigator regarding tolerability and efficacy of the subject.

In general, the target dose should be attained by Days 7-14. For subjects with a body weight ≥45 kg, the target dose range is 120-160 mg/day. A dose of 160 mg/day should not be achieved before Day 14 of treatment, however. For subjects with a body weight <45 kg, the target dose range is 60-80 mg/day. A dose of 80 mg/day should not be achieved before Day 8 of treatment.

After subjects have been titrated to the target dose range, double-blind dosing will continue to the end of Week 4, during which time the dose should be kept stable except that downward adjustments of the dose are allowed if required in the judgment of the investigator for safety and tolerability concerns.

2.2. Sample Size Determination

This study has been designed to have 85% statistical power to show a difference between drug and placebo at conventional levels (ie, 5% level, 2-sided) of statistical significance. Based on the previous Pfizer pediatric study A1281132, the estimated difference in change from baseline of the YMRS total score for ziprasidone versus placebo is –4.55 points (from descriptive and LSMEANs at week 4) with an approximate within-group standard deviation of 8.0.

The effect size used in the design of this study is 0.57, and is consistent with the weighted estimate of the effect size (0.65) of second generation antipsychotics in pediatric bipolar patients published in the recent meta-analysis of Correll et al, which was based on the results of multiple placebo-controlled trials.

Using EAST (version 5.4) for a two-sample t-test, the sample size needed to detect this difference with 85% power at a two-sided significance level of 5% was determined to be 111 randomized subjects. Assuming a 38% dropout rate, the original sample size for this study was 180 subjects.

The randomization will be stratified by weight group (weight <45 kg, weight ≥45 kg) to ensure treatment balance within each of the two weight groups.

2.2.1. Re-estimation of Sample Size

During the conduct of the study, enrollment and subject disposition was monitored on a regular basis as part of the blinded data reviews prepared for presentation to the external Data Monitoring Committee. These reviews have revealed that the actual observed dropout rate in the study is substantially less than the 38% drop-out rate that was assumed for the original sample size estimation procedure.

Based on these observations, the required sample size will be re-estimated by updating the original sample size dropout rate assumption to reflect the actual observed dropout rate, while maintaining all other power assumptions the same.

The re-estimation will require the study to continue to have a minimum of 85% statistical power to show a difference between drug and placebo at conventional levels (ie, 5% level, 2-sided) of statistical significance. The estimated difference in change from baseline of the YMRS total score for ziprasidone versus placebo is –4.55 points with an approximate within-group standard deviation of 8.0. The sample size needed to detect this difference with 85% power at a two-sided significance level of 5% is 111 randomized subjects. Adjusting for dropouts (approximately 20% at week 4, based on the actual observed study drop-out rate), the total sample necessary is approximately 138 subjects (1:1 enrollment with 69 on ziprasidone and 69 on placebo).

Shortly after enrollment into this study commenced, Pfizer was informed by the rater training vendor (on 29 September 2014) that an internal audit had revealed an error in their internal processes. This error resulted in some sites being notified that they could start screening subjects before all raters at the sites had completed all of the required rater training activities. The primary endpoint assessments (YMRS) and/or the diagnostic interviews (KSADS) of 14 subjects were impacted by the rater training issue.

In agreement with the FDA to ensure the robustness of the inferential analysis, study committed to *recruit an additional 14 subjects into the study and conduct a sensitivity analysis excluding these 14 subjects to mitigate the effect of the rater training issue that was identified* (see Section 8.2.2).

During the study subject evaluability meeting, four additional subjects were identified as also having been impacted by a rater training issue affecting either the primary endpoint (YMRS) or the diagnostic interview (KSADS).

With the additional 18 subjects enrolled to mitigate the impact of the rater training issue, and the modified assumption on the dropout rate, the total planned sample for this study **is 156 subjects (1:1 enrollment with 78 on ziprasidone and 78 on placebo)**. (See Table 1).

Table 1. Sample Size Summary

Attribute	Original	Re-Estimation
Power	85%	85%
Difference	-4.55	-4.55
Standard Deviation	8.0	8.0
Alpha	0.05	0.05
Sides	2-sided	2-sided
N Required	111	111
Dropout Rate	38%	20%
Total N	180	138
Adjustment for Rater Training Situation		
Number	14	18
Total w/Rater Adj.	194	**156**

2.3. Study Objectives

2.3.1. Primary Objectives

1. *To assess the efficacy of oral ziprasidone compared with placebo in the treatment of children and adolescents aged 10- 17 with Bipolar I Disorder (current or most recent episode manic) as measured by the change from baseline to Week 4 in the Young Mania Rating Scale (YMRS) total score.*

2. *To evaluate the safety and tolerability of oral ziprasidone over 4 weeks in the treatment of children and adolescents with Bipolar I Disorder (current or most recent episode manic).*

2.3.2. Secondary Objectives

The key secondary objective of the study is to evaluate the efficacy of oral ziprasidone as compared with placebo in the treatment of children and adolescents with Bipolar I Disorder (current or most recent episode manic) as measured by the change from baseline in the Clinical Global Impression of Severity (CGI-S) score.

Other secondary objectives of the study include:

- *An evaluation of the efficacy of oral ziprasidone as compared with placebo as measured by the Clinical Global Impression of Improvement (CGI-I) score.*

CCI [redacted]

[redacted]

- [redacted]
- [redacted]

3. INTERIM ANALYSES, FINAL ANALYSES AND UNBLINDING

No interim efficacy analysis is planned for this study.

This study will utilize an External Data Monitoring Committee (E-DMC). The EDMC will be responsible for ongoing monitoring of the safety of subjects in the study according to the Charter. The EDMC is external to the study team and no unblinded material will be available to study team personnel.

4. HYPOTHESES AND DECISION RULES

4.1. Statistical Hypotheses

The primary null hypothesis: There is no difference in change from baseline in the YMRS total score at Week 4 for subjects randomized to ziprasidone compared to that for subjects randomized to placebo. The corresponding alternative hypothesis is that there is a treatment effect and there is a difference in change from baseline in the YMRS total score at Week 4 for subjects randomized to ziprasidone compared to that for subjects randomized to placebo. Similarly, the null hypothesis for each of the remaining endpoints tested statistically is that the results observed for ziprasidone are no different from those for the placebo group. The corresponding alternative hypothesis is that there is a treatment effect and the ziprasidone treatment group is different from the placebo group.

4.2. Statistical Decision Rules

All hypothesis testing will be conducted using two-sided tests with alpha = 0.05 level of significance. Statistical significance of the key-secondary analysis is dependent on first achieving statistically significant results in the primary analysis. No adjustments for Type I error (at 0.05) will be made for multiple comparisons.

If the test result (from the primary efficacy analysis) is significant (p-value ≤0.05) and in favor of ziprasidone, then efficacy of flexibly-dosed oral ziprasidone in the treatment of children and adolescents with Bipolar I Disorder (current or most recent episode manic) will be established.

5. ANALYSIS SETS

The following sets are defined for use in the analyses:

- Intent-To-Treat (ITT) Analysis Set.
- Per-Protocol (PP) Analysis Set.
- Safety Analysis Set.

Both the ITT and PP analysis sets will be used in the analyses of all efficacy endpoints, with the ITT being primary. The Safety Analysis Set will be used in the analyses of the safety data. *Only the ITT analysis set will be used in the analyses of the* CCI *outcome measure and special safety assessments (CSSRS data, CDRS, and movement disorder scales).*

5.1. ITT Analysis Set

For this protocol, an intent-to-treat (ITT) analysis set will be defined as the set of all patients who were randomized, had baseline measurements, took at least 1 dose of study medication (Ziprasidone or placebo), and with at least 1 post-baseline visit. This definition of ITT may be considered a modified ITT definition; however, it will be referred to as ITT throughout the protocol and the SAP.

5.2. 'Per Protocol' Analysis Set

The per-protocol (PP) analysis set will include all patients in the ITT set without major protocol violations considered to impact the interpretation of the primary efficacy endpoint. Protocol deviations will be reviewed to generate the list of subjects with significant deviations to be excluded from the PP analysis set. The PP exclusion criteria will be finalized prior to breaking the blind.

5.3. Safety Analysis Set

The safety analysis set will include all patients who were randomized and took at least 1 dose of study medication (ziprasidone or placebo).

5.4. Treatment Misallocations

If a subject was:

- Randomized but not treated, these subjects are by definition excluded from the efficacy and safety analyses as all analysis sets require subjects to receive at least 1 dose of study medication.
- Treated but not randomized, these subjects are by definition excluded from the efficacy and safety analyses since randomized treatment is missing.
- Randomized but took incorrect treatment, then they will be reported under their randomized treatment group for all efficacy analyses but will reported under the treatment they received for all safety analyses.

5.5. Protocol Deviations

The following describes any protocol deviations that relate to the PP analysis set stated in Section 5.2:

5.5.1. Deviations Assessed Prior to Randomization

A protocol deviator is a subject who was wrongly enrolled into the study, when inclusion or exclusion criteria were not appropriately satisfied. See protocol sections 4.1 and 4.2.

5.5.2. Deviations Assessed Post-Randomization

Protocol deviations include but are not limited to the following:

- Violations on concomitant medications (See protocol section 5.5);
- Lack of compliance (level to be defined case by case);
- Dosing errors (eg, treatment misallocation, overdose/under-dose).

A full list of protocol deviations will be compiled for the study report prior to database closure. All deviations will be reviewed, and a determination made for the handling of each case. This review will be documented in a note to the study file prior to un-blinding the study database.

6. ENDPOINTS AND COVARIATES

For summaries and analyses of change from baseline, baseline is defined as the pre-dose measurements for variables that have assessment at the Baseline Visit (Day 1), and it is defined as the most recent assessment prior to randomization for variables that have no assessment at the Baseline Visit. For the electrocardiogram (ECG) data, baseline will be defined as the mean of the pre-dose triplicates at the Baseline Visit.

For collection schedules for all endpoints, please refer to the protocol.

6.1. Efficacy Endpoint(s)

6.1.1. Primary Efficacy Endpoint:

- Change from baseline to Week 4 in Young Mania Rating Scale (YMRS) total score.

6.1.2. Key Secondary Efficacy Endpoint:

- Change from baseline to Week 4 in Clinical Global Impression of Severity (CGI-S) score.

6.1.3. Secondary Endpoints:

- Change from baseline to Weeks 1, 2 and 3 in the YMRS total score;
- Change from baseline to Weeks 1, 2 and 3 in the CGI-S score;

- CGI-I scores at Weeks 1, 2, 3, and 4.

CCI [redacted]

[redacted]

[redacted]

6.2. Safety Endpoints

6.2.1. Safety Assessments

- Adverse event monitoring;
- Concomitant treatment monitoring;
- Clinical laboratory monitoring;
- Physical Examinations;
- Blood Pressure and Pulse;
- Height, weight, BMI, BMI z-score, and waist circumference;
- 12-lead ECG.

6.2.2. Special Safety Assessments

- Suicidality assessment monitoring via the Columbia-Suicide Severity Rating Scale (C-SSRS);
- Child Depression Rating Scale - Revised (CDRS-R) monitoring;
- Movement disorder scale (SARS, BAS, and AIMS) monitoring.

6.3. Covariates

Baseline values and weight category will be included as a covariate in the statistical models used to analyze endpoints requiring analysis of covariance (ANCOVA) by the protocol. See Section 8 for details.

7. HANDLING OF MISSING VALUES

For all analyses using mixed model repeated measures models no explicit imputation of missing values will be performed. No imputation techniques (for example: LOCF) will be used for analyzing the endpoint of this study.

8. STATISTICAL METHODOLOGY AND STATISTICAL ANALYSES

8.1. Statistical Methods

8.1.1. Mixed Model Repeated Measures Analyses (MMRM)

The endpoints including the primary, key-secondary, and secondary endpoints will be analyzed *using a mixed model repeated measures analysis of covariance (MMRM); treatment, visit, visit-by-treatment interaction, and weight category as fixed effects and baseline score (with the exception of CGI-I score) as a covariate. Subject effect will enter the MMRM model as a random effect. The estimation method used will be restricted maximum likelihood. An unstructured covariance matrix will be used in the REPEATED statement. In case SAS PROC MIXED fails to converge, other covariance matrix structures will be considered in the sequence of Toeplitz, First-Order Autoregressive, and Variance Components. The EMPIRICAL option will be specified to compute the estimated variance-covariance matrix of the fixed-effects parameters. Type III sums of squares will be used to test both main effects and interactions.*

8.2. Statistical Analyses

8.2.1. Primary Analysis

The primary efficacy endpoint is change from baseline in the YMRS score. The primary time point is Week 4. The primary endpoint will be analyzed using the MMRM model described in Section 8.1.1. *The primary comparison will be between ziprasidone and placebo at Week 4, conducted as a two-sided test at 5% level of significance. Based on the specified model, the point estimate and 95% confidence interval (CI) for the difference in means between the two treatments will be constructed using the least squares means and appropriate standard errors.*

If the test result is significant (p-value $\leq$0.05) *and in favor of ziprasidone, then efficacy of flexibly-dosed oral ziprasidone in the treatment of children and adolescents with Bipolar I Disorder (current or most recent episode manic) will be established.*

A sensitivity analysis will be done by adding an interaction term of treatment and weight category to the primary analysis model specified in Section 8.1.1.

Supplemental analyses of the primary variable will be performed to support the robustness of the conclusions drawn from the primary mixed models repeated measures analysis described above. These supplemental analyses will include the following:

- *Summaries of reasons for discontinuations by treatment group;*
- *Pattern-mixture ANCOVA analysis of change from baseline YMRS scores.*

For the pattern-mixture analysis, grouping of the subjects on the basis of their dropout or missing-data patterns will be explored. 'Patterns' will be defined under the following two cases:

Case 1:

Pattern 1A - all subjects who have completed the study

Pattern 1B - all subjects who have discontinued the study

Case 2:

Pattern 2A - all subjects who have YMRS score for Week 3 or beyond.

Pattern 2B - all subjects who do not have data beyond Week 2.

A mixed effects repeated measures analysis of covariance model will be used to analyze change from baseline in YMRS score under each of the above cases. These models will include terms for treatment, pattern (as defined above), weight category, treatment-by-pattern interaction, visit, treatment by visit interaction, and visit by pattern interaction as fixed effects, baseline as a covariate, and subject as random effect. The estimation method used will be restricted maximum likelihood. The covariance structure among repeated measures will be assumed to be adequately modeled using an unstructured matrix. In case SAS PROC MIXED fails to converge, other covariance matrix structures will be considered in the sequence of Toeplitz, First-Order Autoregressive, and Variance Components. The EMPIRICAL option (in SAS) will be specified to compute the estimated variance-covariance matrix of the fixed-effects parameters. Type III sums of squares will be used to test both main effects and interactions. The primary comparison will be between ziprasidone and placebo at Week 4, averaged over the missing data patterns, and conducted as a two-sided test at 5% level of significance. Based on the specified model, the point estimate and 95% confidence interval (CI) for the difference in means between the two treatments will be constructed using the least squares means and appropriate standard errors.

The percentage of subjects showing a ≥50% reduction in YMRS scores from baseline to Week 4 will be presented descriptively for each group. Additionally, descriptive statistics for change from baseline in YMRS score will be provided by treatment group and visit classification.

8.2.2. Sensitivity Analysis of the primary endpoint

The YMRS and/or KSADS data of 18 subjects were impacted by the rating training issue. A sensitivity analysis of the treatment effect on the primary endpoint, excluding these 18 subjects from the analysis, will use the MMRM analysis specified in the Section 8.1.1. This analysis will be conducted on the ITT population only.

8.2.3. Secondary Analyses

The secondary efficacy endpoints include:

- *Change from baseline to week 4 in the CGI-S score (key secondary endpoint);*

- *Change from baseline to weeks 1, 2 and 3 in YMRS and CGI-S;*
- *Raw CGI-I score at weeks 1, 2, 3 and 4 (as opposed to change from baseline).*

These endpoints will be analyzed using the MMRM model described in Section 8.1.1. The point estimates and 95% CIs for the difference in means between the two treatments will be constructed using the least squares means and appropriate standard errors.

Descriptive statistics for raw CGI-I scores and for change from baseline CGI-S scores will be provided by treatment group and visit classification.

CCI [redacted]

8.2.5. Safety Analyses

The safety assessments include:

- *Adverse Event Reporting;*
- *Concomitant treatment*
- *Clinical Laboratory Testing;*
- *Physical Examinations;*
- *Blood Pressure and Pulse;*
- *Height and Weight, BMI,* BMI z-score, and *waist circumference;*
- *Electrocardiogram.*

A 3-tier analysis approach described below will be used to analyze adverse events. No formal statistical analysis will be conducted on any of the other safety data listed above.

Data for height, weight, BMI, and waist circumference will be summarized by timepoint using descriptive statistics. Additionally, height, weight, and BMI will be standardized using CDC growth charts and the resulting z-scores presented in listings and frequency tables in 1-unit intervals. In addition, frequency tables of above/below a 1-point increase in the z-scores for BMI will be presented.

All randomized subjects who receive at least one dose of study drug will be included in the safety analysis. All adverse events that are observed from the time of first dosing with study medication (at randomization) until the end of study participation will be included in the safety analysis. Adverse events that occurred during treatment with the antipsychotic medication will be reported separately if the event occurred prior to randomization.

All adverse events will be coded according to the Medical Dictionary for Regulatory Activities (MedDRA) and will be summarized by treatment group. The incidence of treatment-emergent adverse events will be tabulated by treatment group and by system organ class. In addition, the incidence of serious adverse events and adverse events that cause withdrawal will be tabulated. All adverse events will be listed.

The following 3-tier approach will be used to summarize AEs. Under this approach, AEs are classified into 1 of 3 tiers. Different analyses will be performed for different tiers and will be documented in the statistical analysis plan as follows:

Tier-1 events: These are pre-specified events of clinical importance and are maintained in a list in the product's Safety Review Plan. This list may be updated, as more is understood about the drug.

Tier-2 events: These are events that are not tier-1 but are "common". A MedDRA Preferred Term (PT) is defined as a tier-2 event if there are at least 4 in any treatment group.

Tier-3 events: These are events that are neither tier-1 nor tier-2 events.

The analysis of adverse events under the 3-tier approach is considered exploratory. There will be no adjustment for multiple comparisons or stratification factors in the analyses. For tier-1 and tier-2 events, the proportion of AEs observed in each treatment groups will be presented along with the point estimates and associated 95% confidence intervals of the risk difference for the ziprasidone group compared with placebo.

For tier-1 events p-values will be included in the presentations. AEs will be arranged in the output sorted in descending point estimate of the risk difference within system organ class. Footnotes in the outputs will include the methods used to derive any p-values and confidence intervals as per Pfizer standards. The tier-1 AEs will be analyzed using the approach of Chan and Zhang (1999)[1] who inverted two one-sided tests at half the significance level each for calculating P-values and confidence intervals. The list of tier-1 AEs is maintained in the Ziprasidone Safety Review Plan.

For tier-2 AEs, both proportion and 95% CIs will be generated using an asymptotic approach (Proc Binomial). A MedDRA PT is defined as a tier-2 event if there are at least 4 in any treatment group. A cross-industry expert team on safety planning, evaluation and reporting

(Crowe et al, 2009)[3] suggests using the "Rule of 4" to define tier-2 events. The "Rule of 4" says that if a trial has 400 or fewer patients per group and there are 4 or more subjects with a given MedDRA PT in any treatment group, then that PT will be categorized as a tier-2 event.

For tier-3 events, simple proportions will be presented.

All clinical laboratory data will be subjected to clinical review, summarized by frequency of events and mean changes from baseline.

All vital sign measurements will be displayed in listings by subject for each sample collection date and time. The measurement taken immediately prior to randomization will be used as the baseline for calculating changes in vital signs.

Centrally over-read ECG variables will be summarized by mean change from baseline to each measurement time for heart rate, PR interval, QRS width, QT interval and QTcF (Fridericia correction) values. Baseline will be defined as the mean of the pre-dose triplicates at the Baseline Visit. Additionally, the incidence of categorical increases in QTc intervals will be provided. Categories for QTcF are ≥450 msec, ≥480 msec, and ≥500 msec. Categories for QTcF as change from baseline are ≥30 msec increase and ≥60 msec increase. QTcF is considered the primary QTc value for measurements of change and for clinical decision making as this correction is more accurate with changes in heart rate. To enable historical comparisons, categorical changes in QTcB (Bazett's correction) also will be tabulated but clinical decisions and interpretation of data will be based on QTcF values.

The frequency of prior and concomitant medications will be summarized by treatment based on the WHO-drug coding dictionary.

8.2.6. Analyses of Special Safety Endpoints

The special safety assessments include:

- *CDRS-R;*
- *C-SSRS;*
- *Movement Disorder Scales (SARS, BAS, and AIMS).*

For special safety assessments, descriptive statistics (n, mean, standard deviation, min, max, 95% CI) by treatment group and visit for the change from baseline will be provided.

For CDRS-R, a listing by treatment group and visit will be generated to include all the CRDS-R ratings for any subject who at any visit had a CDRS-R total score of 40 or more. These are subjects with the potential of developing depression.

For all of the Movement Disorder Scales, descriptive statistics by treatment group and visit for change from baseline will be provided. For the SARS, the total score from the sum of all 10 sub-items will be summarized. For the BAS, only the global clinical assessment item will be summarized. For the AIMS, only the sum of the first seven items (ie, movement cluster) will be summarized. In addition, the same change from baseline MMRM model used for the primary and secondary endpoints will be applied to the Movement Disorder Scales.

The prospectively collected C-SSRS item responses will be mapped to the C-CASA suicidality event codes. No formal statistical hypothesis testing will be undertaken. Listings of both the C-CASA categories as well as the underlying C-SSRS scale data will be prepared. In addition, a summary table of C-CASA category frequencies at screening, baseline, and all post-baseline visits without regard to baseline will be compiled. Additional tables displaying post-baseline worsening and new-onset of suicidality within reporting categories (suicidal ideation) will also be prepared.

8.2.7. Summary of Efficacy Analyses

Table 2. Summary of Efficacy Analyses

Endpoint	Analysis Set	Statistical Method	Objective
Change from baseline in YMRS total score at Week 4	ITT	MMRM	Primary Analysis
Change from baseline in YMRS total score at Week 4	PP	MMRM	Supportive of Primary Analysis
Change from baseline in YMRS total score at Week 4 – Sensitivity analysis (without 18 subjects)	ITT	MMRM	Sensitivity analysis of primary endpoint
Change from baseline in YMRS total score at Week 4 with interaction term of treatment and weight	ITT	MMRM	Sensitivity analysis of primary endpoint
Change from baseline in YMRS total score at Week 4 – Pattern Mixture Analyses	ITT	MMRM	Supportive of Primary Analysis
Change from baseline in YMRS total score at Week 4 – Pattern Mixture Analyses	PP	MMRM	Supportive of Primary Analysis
Change from baseline to week 4 in the CGI-S score	ITT	MMRM	Key-Secondary Analysis
Change from baseline to week 4 in the CGI-S score	PP	MMRM	Supportive of Key-Secondary Analysis
Change from baseline in YMRS total score and CGI-S score at Weeks 1, 2 and 3	ITT	MMRM	Secondary Analysis
Change from baseline in YMRS total score and CGI-S score at Weeks 1, 2 and 3	PP	MMRM	Supportive of Secondary Analysis
CGI-I score at weeks 1, 2, 3 and 4	ITT	MMRM	Secondary Analysis
CGI-I score at weeks 1, 2, 3 and 4	PP	MMRM	Supportive of Secondary Analysis
CCI [redacted]	[redacted]	[redacted]	[redacted]

Descriptive summaries are not included in this summary.

9. REFERENCES

1. Chan ISF, Zhang Z. (1999). Test-based exact confidence intervals for the difference of two binomial proportions. Biometrics, 55:1201–1209.

2. Correll CU, Sheridan EM, et al. (2010). Antipsychotic and mood stabilizer efficacy and tolerability in pediatric and adult patients with bipolar I mania: a comparative analysis of acute, randomized, placebo-controlled trials. Bipolar Disorders, 12:116-141.

3. Crowe B, Xia HA, Berlin JA, et al. (2009). Recommendations for safety planning, data collection, evaluation and reporting during drug, biologic and vaccine development: a report of the safety planning, evaluation and reporting team. Clinical Trials, 6:430-440.

4. Ameena Isa, MD, Ira Bernstein, PhD, Madhukar Trivedi, et al (2014) Childhood Depression Subscales Using Repeated Sessions on Children's Depression Rating Scale – Revised (CDRS-R) Scores. Journal of Child and Adolescent Psychopharmacology Volume 24, Number 6, 2014 Pp. 318–324 DOI: 10.1089/cap.2013.0127

5. Longitudinal Data Analysis by Donald Hedeker and Robert D Gibbons Copyright @2006 John Wiley & Sons. Inc

10. APPENDICES

Appendix 1. STATISTICAL METHODOLOGY DETAILS

Appendix 1.1. Sample SAS Code for the MMRM Analyses

Sample SAS code to fit the mixed model repeated measures specified in Section 8.1.1 is given below. The response variable chg is the change from baseline at the study visits, Trtpn is the treatment group, avisitn is the assigned visit and blwcat is baseline weight category.

For the variance-covariance matrix: test various structures, start with the UN=unstructured matrix, in the case SAS PROC MIXED fails to converge, other covariance matrix structures will be considered in the sequence of Toeplitz, First-Order Autoregressive, and Variance Components/Covariance structure.

```
PROC MIXED DATA=xxxx EMPIRICAL;
        class trtpn avisitn blwcat usubjid;
        model chg = trtpn avisitn wtcat trtpn*avisitn base/ ddfm=kr2;
        repeated avisitn /subject= usubjid r type=UN rcorr covb;
        lsmeans trtpn*avisitn trtpn/ pdiff CL;
RUN;
ODS OUTPUT Diffs=diff
        LSMEANS  = lsmeans;
Run;
```

CCI [redacted]

Appendix 1.3. Sample SAS Code for the CGI-I Raw Scores MMRM Analyses

Sample SAS code to fit the MMRM model specified in 8.1.1 is given below, except there is no baseline in model. The response variable actual is the actual score at each study visit, Trtpn is the treatment group, avisitn is the assigned visit and blwcat is baseline weight category.

```
PROC MIXED DATA=xxxx EMPIRICAL;
        class trtpn avisitn blwcat usubjid;
        model actual = trtpn avisitn wtcat trtpn*avisitn / ddfm=kr2;
        repeated avisitn /subject= usubjid r type=UN rcorr covb;
        lsmeans trtpn*avisitn trtpn/ pdiff CL;
RUN;
ODS OUTPUT Diffs=diff
        LSMEANS  = lsmeans;
Run;
```

Appendix 1.4. Sample SAS Code for the Pattern Mixture Models Case 1 and Case 2

Description of variables:

The response variable chg is the change from baseline at the study visits, Trtpn is the treatment group, avisitn is the assigned visit and blwcat is baseline weight category.

Case 1: dropout variable defined as binary (0= All subjects who have YMRS score for Week 4, 1 = All subjects who do not have data at Week 4)

avgbase as average baseline value of all patients

pcomp1 as proportion completers in Ziprasidone (Number of completers in Zip/Total number of patients in Zip)

pdisc1 as proportion dropouts in Ziprasidone (Number of dropouts in Zip/Total number of patients in Zip)

pcomp2 as proportion completers in placebo

pdisc2 as proportion dropouts in placebo

pct_compldif as **pcomp1** - **pcomp2**

pct_discdif as **pdisc1** - **pdisc2**

Notes to the model below:

(1) The different estimates statements are used to illustrate how the construction of the last estimate statement ZIPRASIDONE vs PLACEBO Averaged over Patterns is done. Same for the difference of ZIPRASIDONE Averaged over Patterns and PLACEBO Averaged over Patterns.

(2) Case 1: As the dropout subgroup has no records on the last visit week 4, there are no corresponding coefficients, and these are hence ignored in estimate specification (in particular three-way interaction coefficients).

```
PROC MIXED DATA=xxxx EMPIRICAL;
        class avisitn trtpn usubjid blwcat dropout;
        model chg=blwcat base avisitn trtpn avisitn*trtpn
             dropout dropout*avisitn dropout*trtpn avisitn*trtpn*dropout / solution covb;
        repeated avisitn /subject= usubjid r type=UN;

estimate "ZIPRASIDONE in Completers"
intercept 1 base avgbase avisitn 0 0 0 1 trtpn 1 0 dropout 1 0  avisitn*trtpn 0 0 0 0 0 0 1 0
avisitn*dropout 0 0 0 0 0 0 1 0 trtpn*dropout 1 0 0 0 avisitn*trtpn*dropout 0 0 0 0 0 0 0 0 0 0
0 0 1 0 / cl singular=1;

estimate "ZIPRASIDONE in Discontinuations"
intercept 1 base avgbase avisitn 0 0 0 1 trtpn 1 0 dropout 0 1 avisitn*trtpn 0 0 0 0 0 0 1 0
trtpn*dropout 0 1 0 0/ cl singular=1;

estimate "ZIPRASIDONE Averaged over Patterns"
intercept 1 base avgbase avisitn 0 0 0 1 trtpn 1 0 dropout pcomp1 pdisc1 avisitn*trtpn 0 0 0
0 0 0 1 0 avisitn*dropout 0 0 0 0 0 0 pcomp1 0 trtpn*dropout pcomp1 pdisc1 0 0
avisitn*trtpn*dropout 0 0 0 0 0 0 0 0 0 0 0 0 pcomp1 0 0 0/ cl singular=1;

estimate "PLACEBO in Completers"
intercept 1 base avgbase avisitn 0 0 0 1 trtpn 0 1 dropout 1 0 avisitn*trtpn 0 0 0 0 0 0 0 1
avisitn*dropout 0 0 0 0 0 0 1 trtpn*dropout 0 0 1 0 avisitn*trtpn*dropout 0 0 0 0 0 0 0 0 0 0 0
0 0 1 0 0/ cl singular=1;

estimate "PLACEBO in Discontinuations"
intercept 1 base avgbase avisitn 0 0 0 1 trtpn 0 1 dropout 0 1 avisitn*trtpn 0 0 0 0 0 0 0 1
trtpn*dropout 0 0 0 1 / cl singular=1;

estimate "PLACEBO Averaged over Patterns"
intercept 1 base avgbase avisitn 0 0 0 1 trtpn 0 1 dropout pcomp2 pdisc2 avisitn*trtpn 0 0 0
0 0 0 0 0 1 avisitn*dropout 0 0 0 0 0 0 pcomp2 0 trtpn*dropout 0 0 pcomp2 pdisc2
avisitn*trtpn*dropout 0 0 0 0 0 0 0 0 0 0 0 0 0 0 pcomp2 0 0 / cl singular=1;

estimate "ZIPRASIDONE vs PLACEBO Averaged over Patterns"
trtpn 1 -1 dropout pct_compldif pct_discdif avisitn*trtpn 0 0 0 0 0 0 1 -1 avisitn*dropout 0
0 0 0 0 0 pct  compldif 0 trtpn*dropout pcomp1 pdisc1 -pcomp2 -pdisc2
avisitn*trtpn*dropout 0 0 0 0 0 0 0 0 0 0 0 0 pcomp1 -pcomp2 0 0 / cl singular=1;

ods output estimates=estimates1;
```

Case 2: dropout variable defined as binary (0= All subjects who have YMRS score for Week 3 or beyond, 1 = All subjects who do not have data beyond Week 2)

avgbase as average baseline value of all patients
pcomp1 as proportion completers in Ziprasidone (Number of completers in Zip/Total number of patients in Zip)
pdisc1 as proportion dropouts in Ziprasidone (Number of dropouts in Zip/Total number of patients in Zip)
pcomp2 as proportion completers in placebo
pdisc2 as proportion dropouts in placebo
pct_compldif as **pcomp1** - **pcomp2**
pct_discdif as **pdisc1** - **pdisc2**

Notes to the model below:

(1) The different estimates statements are used to illustrate how the construction of the last estimate statement ZIPRASIDONE vs PLACEBO Averaged over Patterns is done. Same for the difference of ZIPRASIDONE Averaged over Patterns and PLACEBO Averaged over Patterns.

Case 2: This code is very similar to case 1, except that the proportion of dropouts and completers will be different, also there will not be records for the last 2 visits in the dropout group, the week 3 related coefficients for dropouts will also not be available.

```
PROC MIXED DATA=xxxx EMPIRICAL;
        class avisitn trtpn usubjid blwcat dropout;
        model chg=blwcat base avisitn trtpn avisitn*trtpn
             dropout dropout*avisitn dropout*trtpn avisitn*trtpn*dropout / solution covb;
        repeated avisitn /subject= usubjid r type=UN;

estimate "ZIPRASIDONE in Completers"
intercept 1 base avgbase avisitn 0 0 0 1 trtpn 1 0 dropout 1 0  avisitn*trtpn 0 0 0 0 0 0 1 0
avisitn*dropout 0 0 0 0 0 1 trtpn*dropout 1 0 0 0 avisitn*trtpn*dropout 0 0 0 0 0 0 0 0 0 1 0
/ cl singular=1;

estimate "ZIPRASIDONE in Discontinuations"
intercept 1 base avgbase avisitn 0 0 0 1 trtpn 1 0 dropout 0 1 avisitn*trtpn 0 0 0 0 0 0 1 0
trtpn*dropout 0 1 0 0/ cl singular=1;

estimate "ZIPRASIDONE Averaged over Patterns"
intercept 1 base avgbase avisitn 0 0 0 1 trtpn 1 0 dropout pcomp1 pdisc1 avisitn*trtpn 0 0 0
0 0 0 1 0 avisitn*dropout 0 0 0 0 0 pcomp1 trtpn*dropout pcomp1 pdisc1 0 0
avisitn*trtpn*dropout 0 0 0 0 0 0 0 0 0 0 pcomp1 0/ cl singular=1;

estimate "PLACEBO in Completers"
```

intercept **1** base **avgbase** avisitn **0 0 0 1** trtpn **0 1** dropout **1 0** avisitn*trtpn **0 0 0 0 0 0 0 1** avisitn*dropout **0 0 0 0 0 1** trtpn*dropout **0 0 1 0** avisitn*trtpn*dropout **0 0 0 0 0 0 0 0 0 0 1**/ cl singular=**1**;

estimate "PLACEBO in Discontinuations"
intercept **1** base **avgbase** avisitn **0 0 0 1** trtpn **0 1** dropout **0 1** avisitn*trtpn **0 0 0 0 0 0 0 1** trtpn*dropout **0 0 0 1** / cl singular=**1**;

estimate "PLACEBO Averaged over Patterns"
intercept **1** base **avgbase** avisitn **0 0 0 1** trtpn **0 1** dropout **pcomp2 pdisc2** avisitn*trtpn **0 0 0 0 0 0 0 1** avisitn*dropout **0 0 0 0 0 pcomp2** trtpn*dropout **0 0 pcomp2 pdisc2** avisitn*trtpn*dropout **0 0 0 0 0 0 0 0 0 0 0 pcomp2** / cl singular=**1**;

estimate "ZIPRASIDONE vs PLACEBO Averaged over Patterns"
trtpn **1 -1** dropout **pct_compldif pct_discdif** avisitn*trtpn **0 0 0 0 0 0 1 -1** avisitn*dropout **0 0 0 0 0 pct compldif** trtpn*dropout **pcomp1 pdisc1 -pcomp2 -pdisc2** avisitn*trtpn*dropout **0 0 0 0 0 0 0 0 0 pcomp1 -pcomp2** / cl singular=**1**;

ods output estimates=estimates2;

Appendix 2. VISIT WINDOW

For by-visit inferential analyses on YMRS total score, CGI-S, and CGI-I, if a subject discontinued early from the study, the corresponding ET Visit will be re-slotted back to earlier nominal visits according to the duration since first dosing date. Assuming FACTDAT is the first study medication date and COLLDATE is the date of visit to be re-slotted, windowing will follow the below algorithm:

IF -99 <= COLLDATE - FACTDAT + 1 <= 0 THEN VISIT = SCREENING;
IF COLLDATE - FACTDAT + 1 = 1 THEN VISIT = BASELINE;
IF 2 <= COLLDATE - FACTDAT + 1 <= 10 THEN VISIT = WEEK 1;
IF 11 <= COLLDATE - FACTDAT + 1 <= 17 THEN VISIT = WEEK 2;
IF 18 <= COLLDATE - FACTDAT + 1 <= 24 THEN VISIT = WEEK 3;
IF 25 <= COLLDATE - FACTDAT + 1 <= 31 THEN VISIT = WEEK 4;
IF COLLDATE - FACTDAT + 1> 31 THEN VISIT= FU WEEK 5;