Otsuka Pharmaceutical Development & Commercialization, Inc.

Investigational Medicinal Product
Aripiprazole (OPC-14597, Lu AF41155)

A Phase 1b, Open-label, Multiple-dose, Randomized, Parallel arm, Safety, Tolerability, and Pharmacokinetic Trial of Aripiprazole Intramuscular Depot Administered in the Gluteal Muscle in Adult Subjects with Schizophrenia or Bipolar I Disorder

Protocol No. 031-201-00181

IND: 134612

STATISTICAL ANALYSIS PLAN

Version 2.0

Date: 16 Jul 2020

Table of Contents

Protocol 031-201-00181

List of In-text Tables

Protocol 031-201-00181

List of In-text Figures

List of Appendices

1 Introduction

This statistical analysis plan (SAP) describes the statistical methodology and data analysis algorithms and conventions to be applied for statistical analysis and reporting of data collected under the clinical protocol 031-201-00181. The SAP is consistent with latest version of protocol amendment 2 dated on 31 July 2019.

2 Trial Objectives

2.1 Primary Objectives

The primary objectives of this trial are

- to determine the safety and tolerability of multiple-dose administrations of aripiprazole in adult subjects with schizophrenia or bipolar I disorder;
- to establish the similarity of aripiprazole concentrations on the last day of the dosing interval following the final administration of aripiprazole into the gluteal muscle site in adult subjects with schizophrenia or bipolar I disorder;
- to establish the similarity of aripiprazole exposure over the dosing interval following the administration of aripiprazole into the gluteal muscle site in adult subjects with schizophrenia or bipolar I disorder.

2.2 Secondary Objectives

The secondary objectives of this trial are

- to determine the PK of aripiprazole;
- to determine aripiprazole concentrations 7 days (C_7) and 14 days (C_{14}) after the first dose of aripiprazole for subjects enrolled to the robust sampling schedule;
- to obtain information on the efficacy of aripiprazole over the course of 32 weeks.

3 Trial Design

This is a phase 1b, open-label, multiple-dose, randomized, parallel-arm, multicenter (approximately 20 trial sites) trial in adult subjects with schizophrenia or bipolar I disorder. The trial design schematic is shown in Figure 3-1.

Subjects between the ages of 18 to 64 years old, inclusively, are randomized (1:1 randomization) to receive multiple doses of 2 month (2M) long-acting injectable (LAI) 960 mg (4 injections) or aripiprazole IM depot 400 mg (8 injections) over the course of 32 weeks. Aripiprazole 2M LAI 960 mg are administered at 56-day intervals and aripiprazole IM depot 400 mg are administered at 28-day intervals (+/- 2 days). A final visit occurs 56 (+/- 2) days after the last dose of aripiprazole 2M LAI 960 mg or 28 (+/- 2) days after the last dose of aripiprazole IM depot 400 mg. The randomization is

stratified by PK sampling schedules (robust sampling schedule or sparse sampling schedule) and disease type (schizophrenia or bipolar I disorder). If enrollment of the number of subjects required for the robust sampling schedule is completed, the robust sampling trial sites may enroll new subjects to the sparse sampling schedule.

The trial consists of a screening period of up to 30 days, a treatment period of 169 (± 2) days, and a follow-up period of 56 (± 2) days after administration of the final dose for the aripiprazole 2M LAI 960 mg treatment group, or a treatment period of 197 (± 2) days, a follow-up period with 28 (± 2) days after administration of the final dose for the aripiprazole IM depot 400 mg treatment group. Individual participation is approximately 255 (± 2) days.

Approximately 258 subjects (129 per treatment group) are to be enrolled into the trial with the expectation that approximately 170 subjects (85 per treatment group) complete the trial. An interim analysis may be conducted to ensure adequate power of the trial. Based on the possible sample size re-estimation in the proposed interim analysis, fewer subjects may be enrolled. Up to 76 subjects (38 per treatment group) are to be enrolled to the robust sampling schedule which is expected to result in at least 50 completers (25 per treatment group). And about 182 subjects (91 per treatment arm) are to be enrolled to the sparse sampling schedule which is expected to result in at least 120 completers (60 per treatment arm). Based on the possible sample size re-estimation in the proposed interim analysis, fewer subjects may be enrolled.

Protocol 031-201-00181

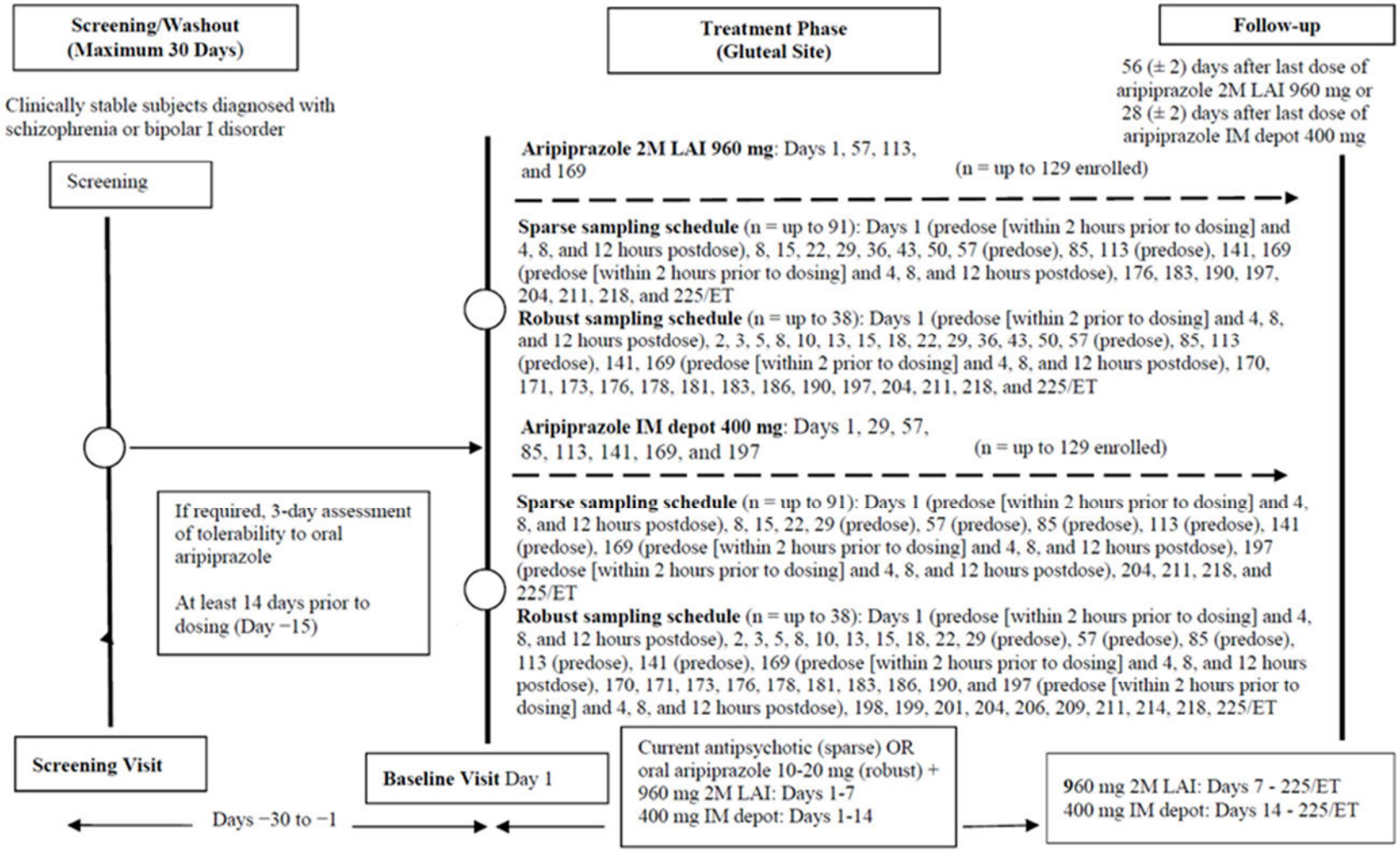


Figure 3-1 Trial Design Schematic

4 Sample Size and Power Justification

To establish the similarity in primary PK variables, the lower bound of the 90% CI of the geometric means ratio (GMR) of C_{56} and AUC_{0-56} after the fourth dose of aripiprazole 2M LAI 960 mg to C_{28} after the eighth dose and the sum of AUC_{0-28} values after the seventh and eighth doses of aripiprazole IM depot 400 mg should be greater than 80%, respectively.

It is estimated that a total of at least 100 subjects (i.e., 50 per group) completing the trial will have at least 80% power to ensure that the lower limit of the 90% CI of the GMR of C_{56} after the fourth dose of aripiprazole 2M LAI 960 mg (test) to C_{28} after the eighth dose of aripiprazole IM depot 400 mg (reference) is greater than 0.80, assuming that the actual GMR of concentrations is 1.0 and the CV is 46%.

Among these 100 subjects, at least 30 completers enrolled to the robust sampling schedule will provide at least 80% power to ensure that the lower limit of the 90% CI of the GMR of AUC_{0-56} of aripiprazole 2M LAI 960 mg (test) to the sum of AUC_{0-28} values of aripiprazole IM depot 400 mg (reference) after the seventh and eighth doses is greater than 0.80, assuming the actual GMR of concentrations is 1.15 and the CV is 40%.

The assumption of CV used in the sample calculation is based on the PK data in the previous multiple-dose aripiprazole IM depot PK Trial 31-12-298.

Assuming a dropout rate of 34%, approximately 152 to 258 subjects are estimated to be enrolled to have 100 to 170 completers based on the proposed interim analysis.

5 Statistical Analysis Datasets

The datasets for analyses are defined as the following:

- Randomized sample consists of all subjects who are randomized.
- Safety sample includes all randomized subjects who receive at least 1 dose of aripiprazole injection, regardless of any protocol violation.
- Pharmacokinetics sample consists of all dosed subjects who have 1 or more evaluable aripiprazole PK parameters. For analysis of the primary PK endpoint, only the completers with available values of the primary endpoints after the last scheduled injection are included.
- Efficacy sample includes all randomized subjects who receive at least 1 dose of aripiprazole injection and have at least 1 efficacy assessment.

5.1 Definition of Baseline and Last Visit

Baseline is defined as the last evaluable assessment prior to aripiprazole IMP injection while last visit is defined as the last evaluable assessment upon completion or early

termination of treatment phase. Only subjects who receive at least 1 dose of investigational medical products and have assessment at both baseline and at least 1 post-baseline visit are included in the analyses of change from baseline.

5.2 Trial Week Windows

The actual assessment date is mapped into the trial week. The definition of trial weeks for primary PK variables are described in Table 5.2-1 and the definitions of trial weeks for efficacy scales and safety scales such as Extrapyramidal Symptoms (EPS) rating scales are included in Table 5.2-2 and Table 5.2-3. Trial days are derived from the formula: Trial Day = Date of assessment - Date of 1st IMP injection + 1. Based on the number of trial days, subjects are mapped to the corresponding trial week. For safety and efficacy variables, only the last observation within the same mapped trial week is used for imputing data on the following trial weeks in the derived last observation carried forward (LOCF) dataset. In addition, only last observations within the same mapped weeks are reported in the summary tables by trial week.

Table 5.2-1 Time Window for Trial Weeks for Primary PK Variables C_{28} and C_{56}

Scheduled Visit (Day)	Mapped Week (Trial Week)	Week Window (Trial Day Range ± 2 days)
197	28 (28 days after the 7th injection of 400 mg Abilify Maintena	195-199
225	32 (28 days after the 8th injection of 400 mg Abilify Maintena or 56 days after the 4th injection of 960 mg aripiprazole 2M LAI)	223-227

Table 5.2-2 Time Window for Mapping Trial Week for Efficacy Variables

Scheduled Visit (Day)	Mapped Week (Trial Week)	Week Window (Trial Day Range)
1	Baseline	≤1 (before first IMP)
29	4	2-43
57	8	44-85
113	16	86-141
169	24	142-183
197	28	184-211
225	32	212-239

Protocol 031-201-00181

Table 5.2-3 Time Window for Mapping Trial Week for Safety Variables

Scheduled Visit (Day)	Mapped Week (Trial Week)	Week Window (Trial Day Range)
1	Baseline	≤1 (before first IMP)
8	1	2-11
15	2	12-18
22	3	19-25
29	4	26-43
57	8	44-71
85	12	72-99
113	16	100-127
141	20	128-155
169	24	156-183
197	28	184-211
225	32	212-239

5.3 Handling of Missing Data

No data imputation is done for missing data in the PK analysis. The last observation carried forward (LOCF) method is used to impute missing data of the safety EPS assessment scales and efficacy assessment scales at postbaseline visits. The observed case (OC) dataset corresponding to a visit consists of data from all subjects who are evaluated at that visit. In the LOCF dataset, missing data at a postbaseline visit are imputed with the value obtained at the nearest preceding visit, except that baseline values are not carried forward to impute missing values at a postbaseline visit. In addition, the Mixed effect Model Repeated Measure (MMRM) for selected efficacy endpoints are explored to address the potentially informational missing (refer to Section 9).

6 Primary and Secondary Outcome Variables

6.1 Primary Safety Variables

Safety and tolerability are based on reported adverse events (AEs), vital signs, ECG parameters, clinical laboratory test results (including blood/serum chemistry, hematology, urinalysis), physical examinations, EPS rating scales (including Simpson-Angus Neurologic Rating Scale [SAS], Abnormal Involuntary Movement Scale [AIMS], and Barnes Akathisia Rating Scale [BARS]), Visual Analog Scale (VAS) scores for pain perception, investigator's assessment of most recent injection site, and suicidality via C-SSRS.

6.2 Primary Pharmacokinetics Variables

The primary PK parameters to be estimated for aripiprazole are as follows:

- Plasma concentration of aripiprazole 56 days postdose (C_{56}) of aripiprazole 2M LAI 960 mg after the fourth dose and plasma concentration of aripiprazole 28 days postdose (C_{28}) of aripiprazole IM depot 400 mg after the eighth dose over the course of 32 weeks;
- $AUC_{0\text{-}56}$ for aripiprazole 2M LAI 960 mg after the fourth dose, and sum of $AUC_{0\text{-}28}$ for aripiprazole IM depot 400 mg after the seventh dose and eighth doses over the course of 32 weeks, based on the PK data from subjects enrolled to the robust sampling schedule.

6.3 Secondary Pharmacokinetics Variables

The following secondary PK parameters are estimated for aripiprazole after the administration of aripiprazole 2M LAI 960 mg and aripiprazole IM depot 400 mg:

- C_{max} and t_{max} after the first and fourth doses of aripiprazole 2M LAI 960 mg;
- $AUC_{0\text{-}56}$ and C_{56} after the first dose of aripiprazole 2M LAI 960 mg;
- $AUC_{0\text{-}28}$ and $AUC_{29\text{-}56}$ after the fourth dose of aripiprazole 2M LAI 960 mg;
- Peak-to-trough percent fluctuation (PTF%) after the fourth dose of aripiprazole 2M LAI 960 mg;
- C_{max} and t_{max} after the first, seventh and eighth doses of aripiprazole IM depot 400 mg;
- $AUC_{0\text{-}28}$ and C_{28} after the first dose of aripiprazole IM depot 400 mg;
- PTF% after the eighth dose of aripiprazole IM depot 400 mg.

C_{max}, t_{max}, AUC, and PTF% are determined only from subjects enrolled to the robust sampling schedule; C_{56} and C_{28} are determined from subjects enrolled to the sparse and robust sampling schedule.

The following PK parameters are estimated for aripiprazole after administration of aripiprazole 2M LAI 960 mg + oral aripiprazole 10 to 20 mg for 7 days and aripiprazole IM depot 400 mg + oral aripiprazole 10 to 20 mg for 14 days (this endpoint is only for subjects enrolled to the robust sampling schedule):

- Plasma concentration of aripiprazole 7 days postdose (C_7)
- Plasma concentration of aripiprazole 14 days postdose (C_{14}).

6.4 Secondary Efficacy Variables

The efficacy of aripiprazole IM depot administration in the gluteal muscle is assessed by the Positive and Negative Syndrome Scale (PANSS; schizophrenia subjects only), Clinical Global Impression Severity (CGI-S; schizophrenia subjects only), Clinical

Global Impression - Improvement (CGI-I), Subjective Well-being under Neuroleptic Treatment-Short Form (SWN-S), Montgomery-Asberg Depression Rating Scale (MADRS; bipolar subjects only), Young Mania Rating Scale (YMRS; bipolar subjects only), and Clinical Global Impression - Bipolar Version (CGI-BP; bipolar subjects only).

7 Summary of Trial Data

7.1 Subject Disposition

Disposition of subjects including screened subjects and/or randomized subjects will be reported by treatment formulation, PK sampling schedule (i.e., robust or sparse) and disease type (i.e., schizophrenia or bipolar I disorder) in addition to summary by center. Durations of subjects staying in the treatment phase will be summarized by week, treatment formulation, PK sampling schedule and disease type. Reasons for discontinuations will be tabulated by treatment formulation, PK sampling schedule and disease type. Subjects who are evaluated at the last scheduled visit (i.e., Day 225 visit for aripiprazole 2M LAI 960 mg and aripiprazole IM depot 400 mg) during the treatment phase will be defined as trial completers.

7.2 Demographic and Baseline Characteristics

Demographic characteristics (including age, weight, height, BMI, sex, race, and ethnicity) will be summarized for randomized sample using descriptive statistics (frequency, mean, median, standard deviation, maximum, minimum, and percentage when applicable) and reported by treatment formulation, PK sampling schedule and disease type.

Disease characteristics at baseline and psychiatric history (including efficacy assessments age of initial diagnosis for schizophrenia or bipolar I disorder) will be presented for the randomized sample by treatment formulation, PK sampling schedule and disease type. In calculating age of first diagnosis for schizophrenia or bipolar disorder, if both month and day are missing, June 1 is used. If only day is missing, date of 1 is used. If only month is missing, month June is used.

7.3 Treatment Compliance

Compliance of trial medication for aripiprazole 2M LAI 960 mg and IM depot 400 mg will be summarized for the randomized sample. Compliance is defined as receiving the aripiprazole IM depot 400 mg injection in the minimum interval of 26 days or 960 mg aripiprazole 2M LAI injection at least 54 days after the prior injection.

Compliance of trial medication of oral aripiprazole in first 2 weeks is derived from total days taking oral tablets divided by 14 days for aripiprazole IM depot 400 mg injection group or 7 days for 960 mg aripiprazole 2M LAI injection group or total days in the trial if patients withdraw from the study before day 15. Proportion of subjects reaching at least 50%, 60%, 70%, 80% and 90% compliance to oral tablets are tabulated by treatment group, disease type for subjects with PK robust sampling schedule.

7.4 Prior and Concomitant Medications

Number and proportion of subjects taking concomitant medications will be tabulated for the randomized sample by treatment formulation and drug classification using the WHO drug dictionary (WHODRUG Global B3, March 2020 or later) and categorized by time periods: prior to start of trial therapy, during the trial therapy, and after the trial therapy.

Proportions of subjects taking a concomitant medication of benzodiazepine will be provided by treatment formulation and/or concomitant medication. Concomitant medications for benzodiazepine are converted into lorazepam equivalent using the appropriate conversion factors. Daily dose of benzodiazepine (lorazepam or equivalent) will be summarized by treatment formulation and/or concomitant medications. The daily dose for each subject is calculated in total dosage of lorazepam or equivalent divided by the duration of IMP treatment. If subjects do not take benzodiazepine, daily dose for benzodiazepine will be considered as 0. Additionally, descriptive statistics for the duration (days) of benzodiazepine will be provided by treatment formulation and/or concomitant medications for subjects taking benzodiazepine only.

7.5 Protocol Deviations

Protocol deviations will be summarized by type of deviations and disease type. Types of deviation include deviations in entry criteria, dosing, randomization, concomitant medication, procedural, etc. In addition, a subject listing will be provided describing the deviations for each subject including protocol deviation due to outbreak of COVID-19.

During the restriction of COVID-19, the robust sampling subjects may remain in clinic until Day 225 and the sparse sampling subjects are asked to come into the clinic on days where they need to receive the once-monthly or once-bimonthly injection of IMP and the final trial visit on Day 225. On those days, all assessments for the sparse sampling subjects will occur as specified in the protocol. On visit days between dosing visits / final study visit, only safety assessments (e.g. AEs, CSSRS, BARS and AIMS) for the sparse sampling subjects will be conducted remotely (virtually) to prioritize the health and safety of these subjects. Therefore, no impact is applicable to primary PK parameters and minimal impact is on secondary PK parameters.

8 Safety Analyses

Safety assessments including adverse events (AEs), vital signs, ECGs, clinical laboratory monitoring (serum chemistry, hematology, urinalysis), physical examinations, EPS, VAS scores for pain perception, investigator's assessment of most recent injection site, and suicidality via the C-SSRS will be summarized for the safety sample by treatment formulation and disease type in addition to all treated subjects.

In summarizing the incidence of clinically relevant abnormalities in clinical laboratory results, vital sign measurements and ECG measurements, a subject must have had an evaluation that meets the specified criteria. Incidence rate will be calculated as the number of subjects having at least 1 abnormality in the treatment phase divided by the number of subjects in the safety sample.

Safety data collected during in-clinic visits and virtual visits are summarized together.

8.1 Exposure to Trial Medication

Exposure to aripiprazole IMP injections for the safety sample will be summarized using percentage of subjects receiving injections and trial day of each injection from date of randomization by treatment formulation, PK sampling schedule and disease type. Exposure to oral aripiprazole during the first 1 or 2 weeks after first injection will be summarized using percentage of subjects exposed to oral aripiprazole and daily dose by week in the randomization phase, treatment formulation, disease type and PK sampling schedule.

8.2 Adverse Events

All AEs are coded using Medical Dictionary for Regulatory Activities (MedDRA Version 23.0 or later). A treatment-emergent adverse event (TEAE) is defined as an AE which starts after start of first IMP injection or an AE continues from baseline of the treatment phase and was serious, trial drug-related or results in death, discontinuation, interruption or reduction of trial medication. The incidences of the following TEAEs will be reported by MedDRA preferred term, treatment formulation and disease type:

- TEAEs by severity,
- Potentially drug related TEAEs,
- TEAEs with an outcome of death,
- Serious TEAEs,
- Discontinuations due to TEAEs,
- Non-Serious TEAEs.

In addition, sub-group analysis for incidence of TEAEs will be summarized by sex (male and female), race (Caucasian, African American and Other), ethnicity (Hispanic/Latino and Not Hispanic/Not Latino), age (< 45 years old and ≥ 45 years old at baseline), BMI (≤ 28 kg/m^2 and > 28 kg/m^2 at baseline), disease type, and PK Robust/Sparse Sampling schedule.

8.3 Vital Signs

Vital signs include body temperature in supine position, systolic/diastolic blood pressure in supine and standing position and heart rate. Potential clinically relevant vital sign abnormalities will be listed, the criteria for which are provided in Appendix 1. Incidences of clinically relevant vital signs abnormalities based on observations during scheduled visits and unscheduled post-baseline visits will be tabulated by treatment formulation and disease type. In addition, vital sign parameters and changes from baseline will be summarized by scheduled visit, treatment formulation and disease type.

If vital sign assessments are repeated for the same trial visit, the last repeat values will be used for production of summary tables. This is accomplished by sorting patient data by visit date and visit time (if applicable) within the same trial visit.

8.4 Physical Examinations

Physical examination data will be listed by subjects.

8.5 Clinical Laboratory

Clinical laboratory tests include serum chemistry, hematology, urinalyses and other lab analyses. Potentially clinically relevant laboratory test abnormalities will be listed by subject and by test. Criteria for potentially clinically relevant laboratory test abnormalities are provided in Appendix 2. Subjects with aspartate aminotransferase (AST) or alanine aminotransferase (ALT) ≥ 3 x upper limit of normal value and total bilirubin ≥ 2 x upper limit of normal value will be listed. The incidences of potentially clinically relevant laboratory tests abnormalities based on observations during scheduled and unscheduled post-baseline visits will be tabulated by treatment formulation and disease type. Summary statistics for clinical laboratory measurements and changes from baseline will be presented by trial visit, treatment formulation and disease type.

If laboratory tests assessments are repeated for the same mapped trial visit, the last repeated values will be used for summary tables. This is accomplished by sorting patient data by visit date and visit time (if applicable) within the same trial visit. If the lab data are recorded as ranges (i.e., including < or > limit of quantification), these data are not

included in the calculations for changes from baseline but included in the calculations for incidences.

The incidence of prolactin level above the upper limit of normal range (potentially clinically relevant change) will be calculated by treatment formulation and disease type. Summary statistics (mean, standard deviation, median, minimum, and maximum) and changes from baseline in prolactin concentration will be presented by trial visit, treatment formulation and disease type. Similar analyses will be performed for male and female, respectively. Additionally, incidences of treatment-emergent relevant changes in lipid parameters and glucose (as defined in Table 8.5-1) will be summarized by treatment formulation and disease type, though those blood samples are not required to be collected at the fasted state.

Table 8.5-1 Criteria for Treatment-emergent Relevant Changes in Lipids Parameters and Glucose

Parameters	Baseline	Any Post-baseline
Total Cholesterol (mg/dL)		
	Normal <200	High >=240
	Borderline 200~<240	High >=240
	Normal/Borderline <240	High >=240
	Normal <200	Borderline/High >=200
	Any Value	Increased >=40
LDL Cholesterol (mg/dL)		
	Normal <100	High >=160
	Borderline 100~<160	High >=160
	Normal/Borderline <160	High >=160
	Normal <100	Borderline/High >=100
	Any Value	Increased >=30
HDL Cholesterol (mg/dL)		
	Normal >=40	Low<40
	Any Value	Decreased >=20
Triglycerides (mg/dL)		
	Normal <150	High >=200
	Normal <150	Very High >=500
	Borderline 150~<200	High >=200
	Borderline 150~<200	Very High >=500
	Normal/Borderline <200	High >=200
	Normal/Borderline <200	Very High >=500
	Normal <150	Borderline/High/Very High >=150
	Normal/Borderline /High >=200	Very High >=500
	Any Value	Increased >=50
Glucose (mg/dL)		
	Normal <100	High >=126
	Impaired Fasting Glucose 100~<126	High >=126

Protocol 031-201-00181

Table 8.5-1	**Criteria for Treatment-emergent Relevant Changes in Lipids Parameters and Glucose**	
Parameters	Baseline	Any Post-baseline
	Normal/ Impaired Fasting Glucose <126	High >=126
	Any Value	Increased >=10

Note: The National Cholesterol Education Program (NCEP) Adult Treatment Program Classifications of lipids (refer to fasting lipid measurements) are used for lipid parameters and the American Diabetes Association Criteria are used for glucose.

8.6 Electrocardiogram

ECG measurements including PR, QRS, QT and RR are taken from three consecutive complexes (representing three consecutive heart beats). The mean value of the available ECG measurements will be calculated and used in the summary table. For the calculation of QT correction by heart rate, QTc is considered as missing only if all three consecutive beats are unreadable. For each ECG, QT and RR intervals from three consecutive complexes (representing three consecutive heart beats) will be measured manually. The QT correction is performed on beat-to-beat basis. The mean of ratios (beat-to-beat) is calculated by (QTc1 + QTc2 + QTc3)/3 using each QT-RR pair: (QT1, RR1), (QT2, RR2) and (QT3, RR3). The corrected QT intervals for QTcB, QTcF, and QTcN are defined as follows:

- QTcB is the length of the QT interval corrected for heart rate by Bazett's formula: $QTcB = QT / (RR)^{1/2}$,
- QTcF is the length of the QT interval corrected for heart rate by Fredericia's formula: $QTcF = QT / (RR)^{1/3}$,
- QTcN is the length of the QT interval corrected for heart rate by the FDA Neuropharm Division formula: $QTcN = QT/ (RR)^{0.37}$.

Potentially clinically relevant ECG abnormalities will be listed by subject. The criteria for potentially clinically relevant ECG abnormalities are provided in Appendix 3. Incidences of abnormal ECGs of potential clinical relevance, based on the observations during scheduled and unscheduled post-baseline visits will be tabulated by treatment formulation and disease type. Descriptive statistics of change from baseline in ECG intervals of PR, QRS, RR, QT and QTc, will be presented by trial visit, treatment formulation and disease type. In addition, incidences of categorical change in ECG - QTc will be provided for the following categories: change from baseline >30 millisecond (MS) to <60 MS, change from baseline ≥60 MS, new onset >450 MS, new onset >480 MS and new onset >500 MS.

If ECG assessments are repeated for the same trial visit, the last repeated values will be used for generation of mean change from baseline. This is accomplished by sorting patient data by visit date and visit time (if applicable) within the same trial visit.

8.7 Suicidality

Suicidality data collected from the C-SSRS assessments will be summarized by treatment formulation and disease type in incidence of suicidality; suicidal behavior only, emergence of suicidal behavior; suicidal ideation only, emergence of suicidal ideation, emergence of serious suicidal ideation, and worsening of suicidal ideation at trial visits and last visit in addition to overall during the treatment phase.

Suicidality is defined as reporting any suicidal ideation or behavior. Suicidal behavior only is defined as reporting any type of suicidal behaviors (actual attempt, interrupted attempt, aborted attempt and preparatory acts or behavior) throughout assessment period. Emergence of suicidal behavior is defined as having no suicidal behavior at baseline and reporting any type of behavior at post-baseline. Suicidal ideation only is defined as reporting any type of suicidal ideation. Emergence of suicidal ideation is defined as having no suicidal ideation at baseline and reporting any type of ideation during treatment. Emergence of serious suicidal ideation is defined as having no suicidal ideation at baseline and reporting serious suicidal ideation with score of 4 or 5 on suicidal ideation severity rating during the treatment. In assessment of suicidal ideation and level of severity, each person is given a score of 0 (no ideation present) to 5 (active ideation with plan and intent). Worsening of suicidal ideation is defined as having greater severity of ideation for the most severe suicidal ideation at post baseline than at baseline. Numbers of patients reporting suicidal behavior are also reported by type of suicidal behavior.

Descriptive statistics of the incidence of suicidality and incidence of suicidality by type (suicidal behavior and suicidal ideation), as well as the mean changes of Suicidal Ideation Intensity (SSI) total scores for most severe ideation from baseline will be reported by trial visit, treatment formulation and disease type. The suicidal ideation intensity total score is the sum of suicidal ideation severity rating scores for frequency, duration, controllability, deterrents and reasons for ideation. For each item, each subject gets intensity score from 0 (none) to 5 (worst). Therefore, SSI total score is a range from 0 to 25. If the subject did not endorse any suicidal ideation, a score of 0 is given for the intensity scale.

There are 9 categories for Columbia Classification Algorithm of Suicide Assessment (C-CASA) method, which are coded as below,

1. completed suicide;

2. suicide attempt;
3. preparatory actions toward imminent suicidal behavior;
4. suicidal ideation;
5. self-injurious behavior, intent unknown;
6. not enough information: death;
7. non-suicidal self-injurious behavior;
8. other (Accident; Psychiatric, Medical);
9. not enough information: non-death.

Data collected from C-SSRS will be mapped into C-CASA category as specified in Table 8.7-1. Incidence of potential suicidality events using C-CASA category will be provided for safety sample.

Table 8.7-1 Mapping C-SSRS to C-CASA

C-SSRS	C-CASA Category
Completed Suicide	1. Completed Suicide
Actual Attempt	2. Suicide Attempt
-Interrupted Attempt -Aborted Attempt -Preparatory Acts or Behavior	3. Preparatory Actions Toward Imminent Suicidal Behavior
1. Wish to Die 2. Non-Specific Active Suicidal Thoughts 3. Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act 4. Active Suicidal Ideation with Some intent To Act, without Specific Plan 5. Active Suicidal Ideation with specific Plan and Intent	4. Suicidal Ideation
Non-suicidal Self-injurious Behavior	7. Non-suicidal Self-injurious Behavior

8.8 Extrapyramidal Symptoms Scales

Extrapyramidal Symptoms (EPS) rating scales include the Simpson-Angus Scale (SAS), Barnes Akathisia Rating Scale (BARS) and Abnormal Involuntary Movement Scale (AIMS). The variables to be analyzed are,

- Change from Baseline in SAS Total Score,
- Change from Baseline in BARS Global Score,
- Change from Baseline in AIMS Movement Rating Score.

The SAS total score (range 10 - 50) is the sum of the rating scores for 10 items from the SAS panel in the Case Report Form (CRF). The BARS global score (range 0 - 5) is derived from the global clinical assessment of akathisia from the BARS panel in the CRF. The AIMS movement rating score (range 0 - 28) is the sum of the rating scores for facial and oral moments (i.e., item 1 - 4), extremity movements (i.e. item 5 - 6), and trunk movements (i.e. item 7). A missing value of any item for the SAS or the AIMS movements scale could result in a missing SAS total score or AIMS movement rating

score. LOCF missing data imputation will be performed those scores. Descriptive statistics of the EPS rating scales for both LOCF and OC data and the mean changes of the scales from baseline will be reported by trial week, treatment formulation and disease type.

8.9 Injection Site Pain

Descriptive statistics for VAS scores for pain perception will be tabulated by treatment formulation and disease type at each injection (predose and after dose). Investigator's assessment of most recent injection site including pain, swelling, redness, and induration will be reported in 4-point categorical scale (absent, mild, moderate and severe) and will be summarized by treatment formulation and disease type at each injection (predose and after dose).

8.10 Special Interest Adverse Events

Incidence of treatment emergent special interest adverse events will be tabulated by treatment formulation and MedDRA preferred terms for EPS-related AEs, weight-related AEs, glucose-related AEs, lipid-related AEs, white blood cell abnormalities, orthostasis-related AEs, prolactin-related AEs, QT-interval AEs, suicidality -related AEs and injection-site-related AEs. In addition, proportion of subjects received anticholinergic medication for treatment-emergent EPS-related events will be presented.

9 Efficacy Analyses

Efficacy assessments for all subjects includes Clinical Global Impression - Improvement (CGI-I) and Subjective Well-being under Neuroleptic Treatment - Short Form (SWN-S). The Positive and Negative Syndrome Scale (PANSS) and Clinical Global Impression - Severity (CGI-S) are assessed for schizophrenia subjects only. Montgomery-Asberg Depression Rating Scale (MADRS), Young Mania Rating Scale (YMRS) and Clinical Global Impression - Bipolar Version (CGI-BP) are assessed for bipolar I disorder subjects only. All efficacy assessments will be summarized for the efficacy sample at each mapped trial week by treatment formulation and disease type (if applicable).

PANSS total score (range 30-210) is the sum of the rating scores for 7 positive scale items, 7 negative scale items, and 16 general psychopathology scale items from the PANSS scale. A missing value for any item would result in a missing PANSS total score. PANSS positive subscale score (range 7-49) is the sum of the rating scores for the 7 positive scale items from the PANSS scale. PANSS negative subscale score (range 7-49) is the sum of the rating scores for the 7 negative scale items from the PANSS scale. CGI-S score (range 0-7), CGI-I score (range 0-7), CGI-BP severity score (range 0-7) and

CGI-BP change from preceding phase score (range 0-7) are directly retrieved from CRF data.

SWN-S total score (range 20-120) is the sum of the rating scores for 10 positive scale items (items 2+, 3+, 5+, 7+, 8+, 13+, 15+, 18+, 19+, 20+) and 10 negative scale items (items 1-, 4-, 6-, 9-, 10-, 11-, 12-, 14-, 16-, 17-). SWN-S subscale total score (range 4-24) is the sum of the rating scores for the 4 items in each of the 5 subscales: mental functioning (3+, 7+, 11-, 17-), social integration (8+, 13+, 6-, 14-), emotional regulation (18+, 20+, 4-, 10-), physical functioning (2+, 5+, 9-, 16-), and self-control (15+, 19+, 1-, 12-). For items marked with a '+', the scoring is 'not at all = 1', 'hardly at all = 2', 'a little = 3', 'somewhat = 4', 'much = 5', and 'very much = 6'. For items marked with a '-', the scoring is reversed. response choices and scoring are as follows: not at all = 6, hardly at all = 5, a little = 4, somewhat = 3, much = 2, very much = 1. A missing value of any items in the SWN-S would result in a missing total or a subscale total score.

The MADRS is utilized as the primary assessment of a subject's level of depressive symptoms and must be administered using a structured interview guide. This scale consists of 10 items each with 7 defined grades of severity on 0 to 6 scale (reported sadness, apparent sadness, inner tension, reduced sleep, reduced appetite, concentration difficulties, lassitude, inability to feel, pessimistic thoughts, and suicidal thoughts). MADRS total score is sum of 10 item score with range from 0~60. The missing score for any 1 item will result in a missing MADRS total score. MADRS total score will be summarized by trial week for each treatment formulation.

The YMRS consists of 11 items assessing the core symptoms of mania and is based on the subject's subjective report of his or her clinical condition. Additional information is based upon clinical observations made during the clinical interview. Each item has 5 defined categories of severity with 4 items graded on a 0 to 8 scale (irritability, speech, content, and disruptive-aggressive behavior) and 7 items graded on a 0 to 4 scale (elevated mood, increased motor activity/energy, sexual interest, sleep, language-thought disorder, appearance, and insight). YMRS total score is sum of 11 item score with range from 0~60. The missing score for any 1 item will result in a missing YMRS total score. YMRS total score will be summarized by trial week for each treatment formulation.

The CGI-BP scale refers to the global impression of the subject with respect to bipolar disorder. The scale rates the subject's severity of illness (CGI-BP-Severity: mania, depression, and overall bipolar illness) and change from baseline visit (CGI-BP Change from Preceding Phase: mania, depression, and overall bipolar illness) based on a 7-point scale (rang 1-7).

LOCF missing data imputation will be performed for all efficacy assessments. Changes from baseline in SWN-S total score and SWN-S subscale total score for all efficacy sample in addition to PANSS total score, PANSS positive subscale total score, PANSS negative subscale total score, CGI-Severity score, MADRS total score, YMRS total score, CGI-BP-Severity score will be summarized using descriptive statistics, if applicable. CGI-Improvement score and CGI-BP change from preceding phase score are also summarized using descriptive statistics, if applicable.

In addition, the linear mixed-effects mode will be used to explore the effect of the treatment and change from baseline over the time for efficacy endpoints, i.e. SWN-S total score, PANSS total score and YMRS total score. Change from baseline in those efficacy endpoints will be analyzed using MMRM analysis with a restricted maximum likelihood (REML) approach. Analyses will include the categorically fixed effects of treatment, disease type (if applicable), PK sample scheduling (if applicable), trial week, and treatment-by trial week interaction, as well as the covariates of baseline-score-by-week interaction in addition to subject as random effect. An unstructured covariance structure will be used to model the within-subject errors and Kenward-Rodger degree of freedom will be used to test the fixed effects. If this analysis approach fails to converge, the covariance structures of heterogeneous compound symmetry will be utilized in the analysis and the "sandwich" estimator of the variance covariance matrix of the fixed effects parameters will be used in order to deal with possible model misspecification of the covariance matrix. 95% confidence intervals (CIs) for difference of changes from baseline in efficacy endpoints or values of efficacy endpoints between aripiprazole 2M LAI 960mg and IM depot 400 mg will be provided at trail weeks as well.

10 Pharmacokinetics Analysis

The pharmacokinetic parameters will be calculated using Noncompartmental Analysis (NCA) SAS macro (version 1.0). In general, the individual and summary tables, using descriptive statistics (N, median, mean, standard deviation, coefficient of variation, minimum, maximum), of pharmacokinetic parameters will be summarized descriptively by treatment formulation. Descriptive statistics for aripiprazole and dehydro-aripiprazole concentration data will be presented.

10.1 Primary Pharmacokinetics Endpoint Analyses

The following PK parameters will be estimated and summarized descriptively for aripiprazole: C_{56} and AUC_{0-56} of aripiprazole 2M LAI 960 mg after the fourth dose, C_{28} of aripiprazole IM depot 400 mg after the eighth dose, and the sum of AUC_{0-28} of aripiprazole IM depot 400 mg after the seventh and eighth dose over the course of 32

weeks. Values for AUC_{0-28}, and AUC_{0-56} will be summarized based on samples from subjects enrolled to the robust sampling schedule; whereas C_{28} and C_{56} will be summarized based on samples from subjects enrolled to the robust and sparse sampling schedules.

The 90% confidence interval (CI) of GMR of C_{56} of aripiprazole after the fourth injection of aripiprazole 2M LAI 960 mg (test) to C_{28} after the eighth injection of aripiprazole IM depot 400 mg (reference) will be provided. An analysis of variance will be performed on the natural-log transformed PK parameters using the MIXED procedure in the Statistical Analysis System (SAS). The mixed-effects linear model will include treatment formulation, disease type (if applicable), and PK sampling schedule (if applicable) as fixed effects and subject as a random effect. The least squares mean for the 2 treatment formulations, their difference, and the 90% CI for their difference will be derived. The antilog of the confidence limits will provide the 90% CI for the GMR of the 2 treatments. For subjects enrolled to the robust sampling schedule, the values of AUC_{0-28} after the seventh and eighth doses of aripiprazole IM depot 400 mg (reference) will be summed up before analyzing by the same method together with AUC_{0-56} after the fourth injection of aripiprazole 2M LAI 960 mg (test). Only subjects who receive the fourth dose of aripiprazole 2M LAI 960 mg or the seventh and eighth dose of aripiprazole IM depot 400 mg and have C_{56}, C_{28}, AUC_{0-28}, and AUC_{0-56} values determined for the respective treatments will be included in the analysis.

10.2 Secondary Pharmacokinetics Endpoint Analyses

Individual and summary tables, using descriptive statistics (N, median, mean, standard deviation, coefficient of variation, minimum, maximum) will be provided for the following aripiprazole PK endpoints,

- C_{max} and t_{max} after the first and fourth doses of aripiprazole 2M LAI 960 mg,
- AUC_{0-56} and C_{56} after the first dose of aripiprazole 2M LAI 960 mg,
- AUC_{0-28} and AUC_{29-56} after the fourth dose of aripiprazole 2M LAI 960 mg,
- PTF% after the fourth dose of aripiprazole 2M LAI 960 mg (for aripiprazole only),
- C_{max} and t_{max} after the first, seventh and eighth doses of aripiprazole IM depot 400 mg,
- AUC_{0-28} and C_{28} after the first dose of aripiprazole IM depot 400 mg,
- PTF% after the eighth dose of aripiprazole IM depot 400 mg (for aripiprazole only),

- C_7 and C_{14} after the first injection of aripiprazole 2M LAI 960 mg and aripiprazole IM depot 400 mg from subjects enrolled to the robust sampling schedule.

Values for C_{max}, AUC, t_{max}, and PTF% will be summarized based on samples from subjects enrolled to the robust sampling schedule; whereas C_{28} and C_{56} will be summarized based on samples from subjects enrolled to the robust and sparse sampling schedules.

[redacted]

Protocol 031-201-00181

Appendix 1 Criteria for Identifying Vital Sign Values of Potential Clinical Relevance

Variable	Criterion Value[a]	Change Relative to Baseline[a]
Heart Rate[b]	> 120 bpm < 50 bpm	≥ 15 bpm increase ≥ 15 bpm decrease
Systolic Blood Pressure[b]	> 180 mmHg < 90 mmHg	≥ 20 mmHg increase ≥ 20 mmHg decrease
Diastolic Blood Pressure[b]	> 105 mmHg < 50 mmHg	≥ 15 mmHg increase ≥ 15 mmHg decrease
Orthostatic Hypotension	≥ 20 mmHg decrease in systolic blood pressure and a ≥ 25 bpm increase in heart rate from supine to sitting/standing	Not Applicable
Weight	-	≥ 7% increase ≥ 7% decrease
Temperature	≥ 37.8°C	≥ 1.1°C increase

a In order to be identified as potentially clinically relevant, an on-treatment value must meet the "Criterion Value" and also represent a change from the subject's baseline value of at least the magnitude shown in the "Change Relative to Baseline" column.

b As defined in "Supplementary Suggestions for Preparing an Integrated Summary of Safety Information in an Original NDA Submission and for Organizing Information in Periodic Safety Updates," FDA Division of Neuropharmacological Drug Products draft (2/27/87).

Protocol 031-201-00181

Appendix 2 Criteria for Identifying Laboratory Values of Potential Clinical Relevance

Laboratory Tests	Criteria
Chemistry	
AST (SGOT)	≥ 3 x upper limit of normal (ULN)
ALT (SGPT)	≥ 3 x ULN
Alkaline phosphatase	≥ 3 x ULN
LDH	≥ 3 x ULN
BUN	≥ 30 mg/dL
Creatinine	≥ 2.0 mg/dL
Uric acid	
Men	≥ 10.5 mg/dL
Women	≥ 8.5 mg/dL
Bilirubin (total)	≥ 2.0 mg/dL
CPK	≥ 3 x ULN
Prolactin	> ULN
Hematology	
Hematocrit	
Men	≤ 37 % and decrease of ≥ 3 percentage points from baseline
Women	≤ 32 % and decrease of ≥ 3 percentage points from baseline
Hemoglobin	
Men	≤ 11.5 g/dL
Women	≤ 9.5 g/dL
White blood count	≤ 2,800 mm^3 or ≥ 16,000 mm^3
Eosinophils	≥ 10%
Neutrophils	≤ 1,500/mm^3
Platelet count	≤ 75,000/mm^3 or ≥ 700,000/mm^3
Urinalysis [a]	
Protein	Increase of ≥ 2 units
Glucose	Increase of ≥ 2 units
Casts	Increase of ≥ 2 units
Additional Criteria	
Chloride	≤ 90 mEq/L or ≥ 118 mEq/L
Potassium	≤ 3.0 mEq/L or ≥ 5.5 mEq/L
Sodium	≤ 126 mEq/L or ≥ 156 mEq/L
Calcium	≤ 8.2 mg/dL or ≥ 12 mg/dL
Glucose	
Fasting	≥ 115 mg/dL
Non-fasting	≥ 200 mg/dL
Total cholesterol, fasting	≥ 240 mg/dL
LDL cholesterol, fasting	≥ 160 mg/dL
HDL cholesterol, fasting	≤ 30 mg/dL
Triglycerides, fasting	
Men	≥ 160 mg/dL
Women	≥ 120 mg/dL

a As defined in "Supplementary Suggestions for Preparing an Integrated Summary of Safety Information in an Original NDA Submission and for Organizing Information in Periodic Safety Updates," FDA Division of Neuropharmacological Drug Products draft (2/27/87).

Protocol 031-201-00181

Appendix 3 Criteria for Identifying ECG Measurements of Potential Clinical Relevance

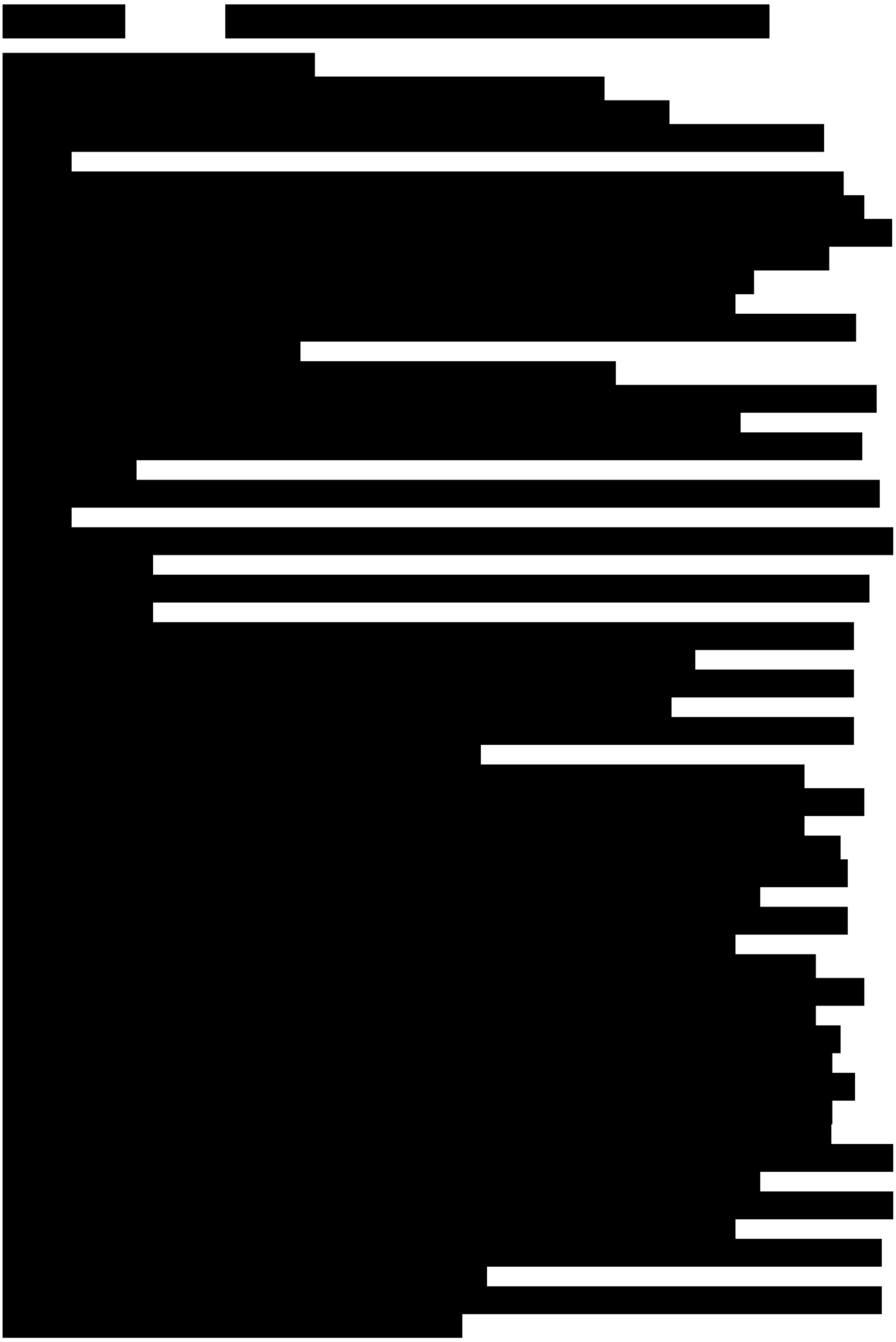
Variable	Criterion Value[a]	Change Relative to Baseline[a]
Rate		
Tachycardia	≥ 120 bpm	increase of ≥ 15 bpm
Bradycardia	≤ 50 bpm	decrease of ≥ 15 bpm
Rhythm		
Sinus tachycardia[b]	≥ 120 bpm	increase of ≥ 15 bpm
Sinus bradycardia[c]	≤ 50 bpm	decrease of ≥ 15 bpm
Supraventricular premature beat	all	not present → present
Ventricular premature beat	all	not present → present
Supraventricular tachycardia	all	not present → present
Ventricular tachycardia	all	not present → present
Atrial fibrillation	all	not present → present
Atrial flutter	all	not present → present
Conduction		
1° atrioventricular block	PR ≥ 0.20 second	increase of ≥ 0.05 second
2° atrioventricular block	all	not present → present
3° atrioventricular block	all	not present → present
Left bundle-branch block	all	not present → present
Right bundle-branch block	all	not present → present
Pre-excitation syndrome	all	not present → present
Other intraventricular conduction block[d]	QRS ≥ 0.12 second	increase of ≥ 0.02 second
Infarction		
Acute or subacute	all	not present → present
Old	all	not present → present ≥ 12 weeks post trial entry
ST/T Morphological		
Myocardial ischemia	all	not present → present
Symmetrical T-wave inversion	all	not present → present
Increase in QTc	QTc > =450 msec	> =10% increase

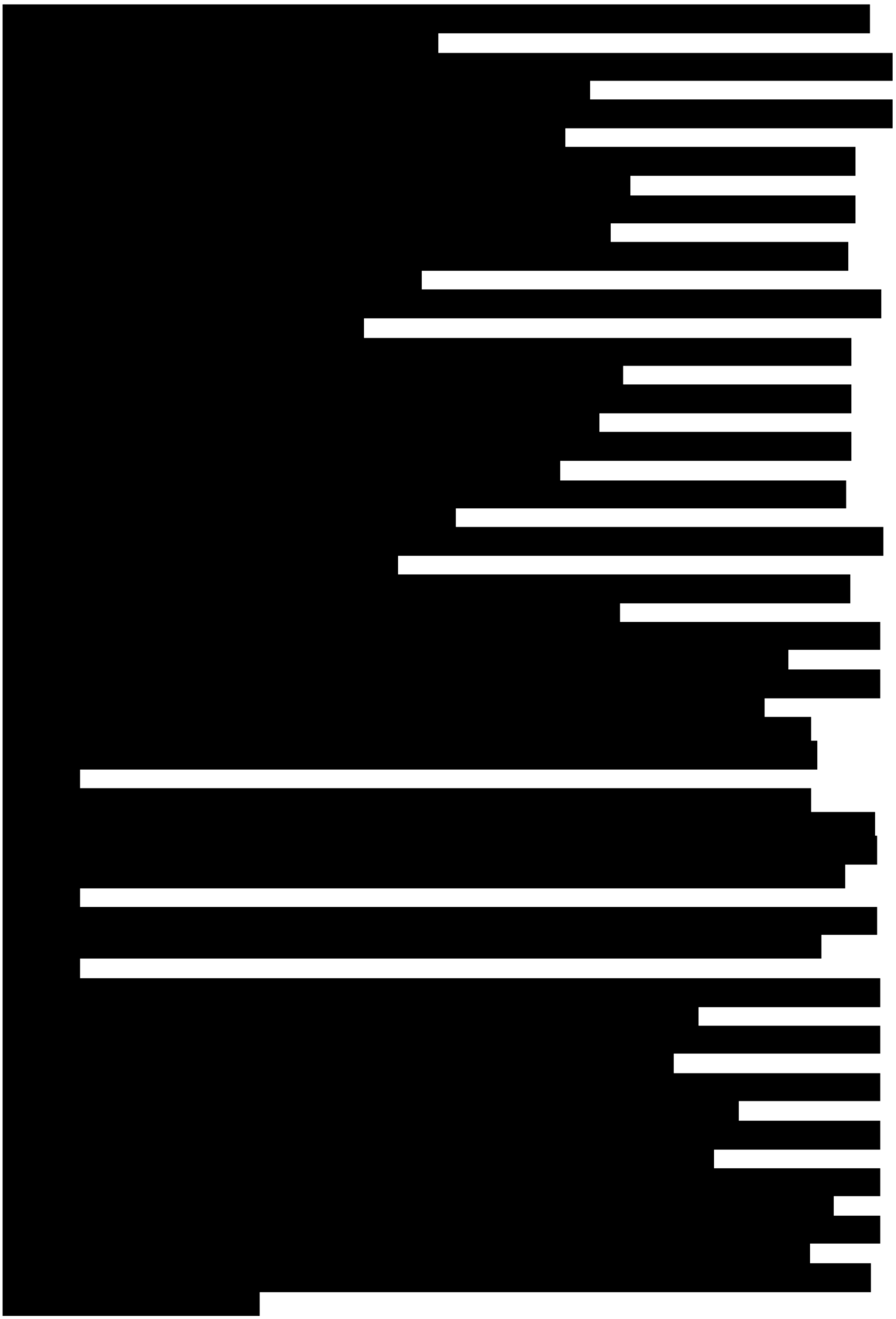
a In order to be identified as potentially clinically relevant, an on-treatment value must meet the "Criterion Value" and also represent a change from the subject's baseline value of at least the magnitude shown in the "Change Relative to Baseline" column.

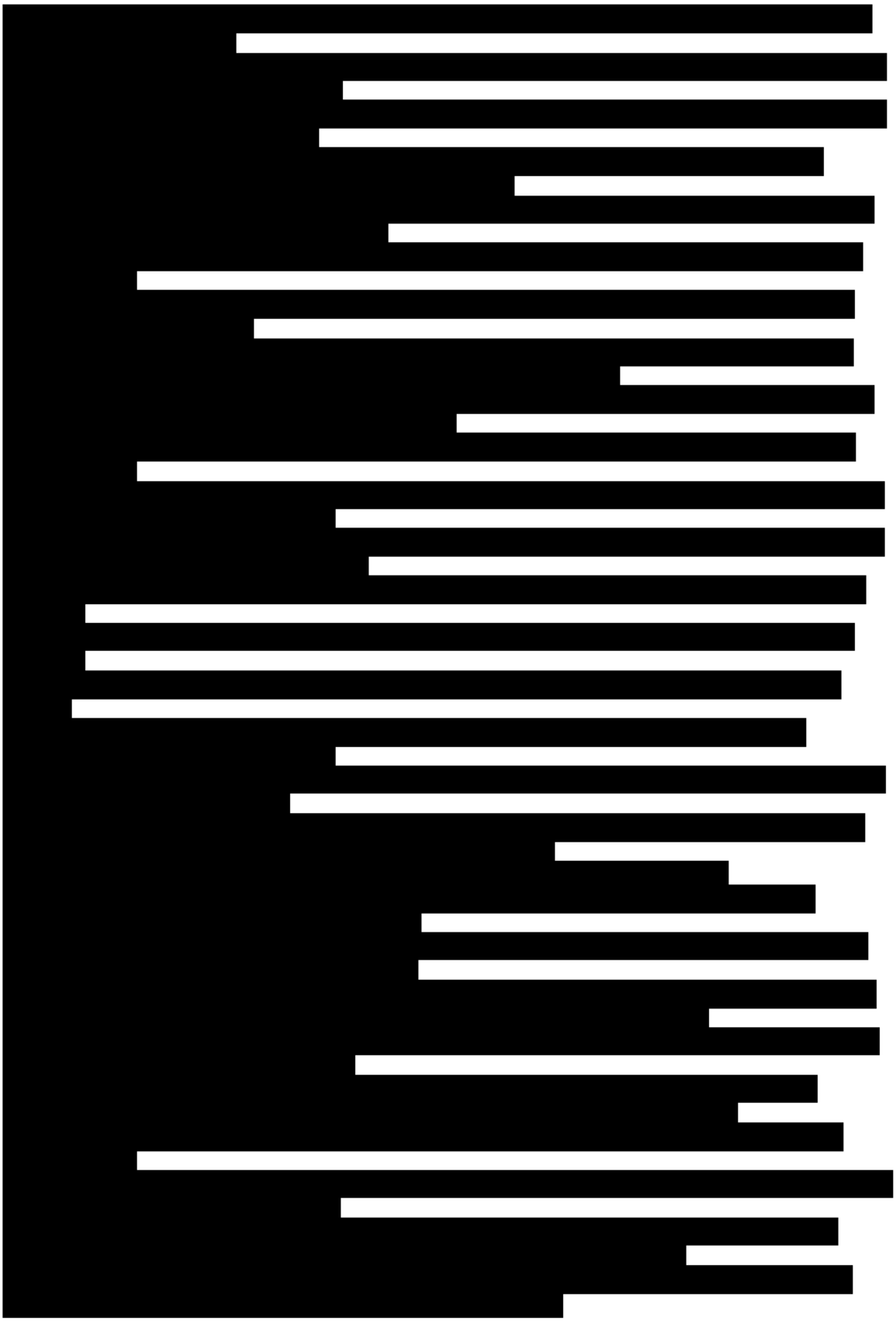
b No current diagnosis of supraventricular tachycardia, ventricular tachycardia, atrial fibrillation, atrial flutter, or other rhythm abnormality.

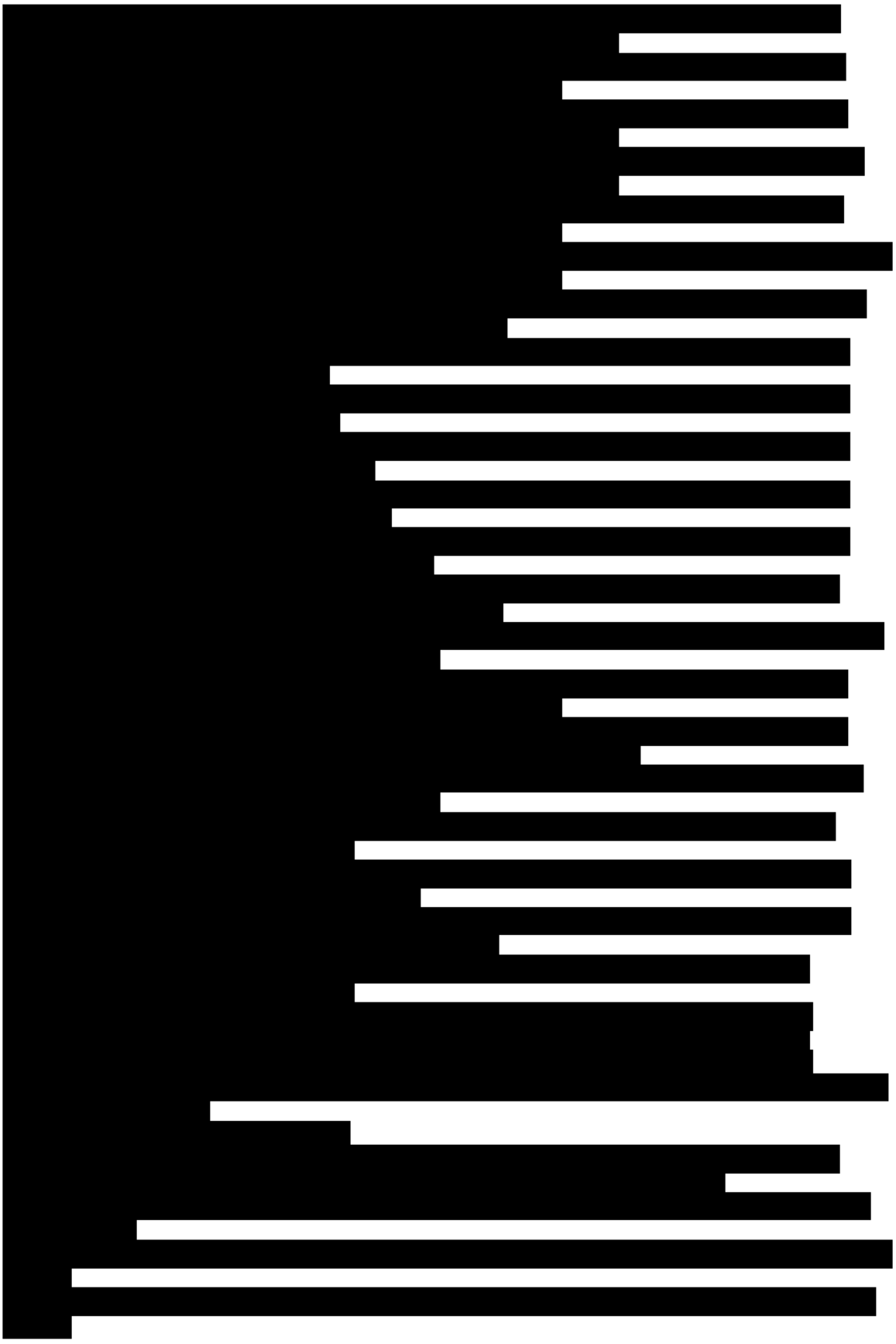
c No current diagnosis of atrial fibrillation, atrial flutter, or other rhythm abnormality.

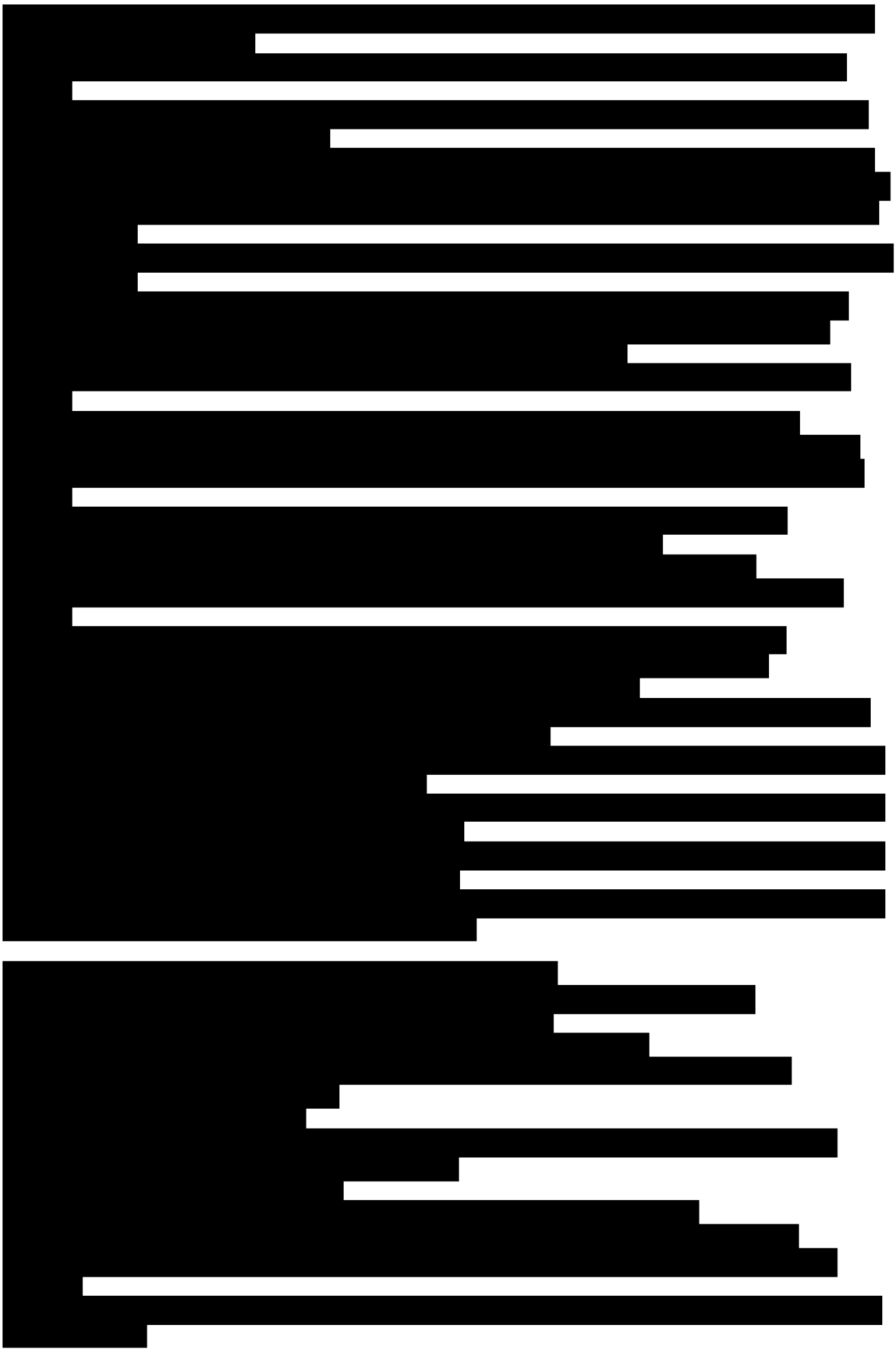
d No current diagnosis of left bundle branch block or right bundle branch block.











[redacted]

This page is a manifestation of an electronically captured signature

SIGNATURE PAGE

Document Name: P03120100181_Statistical Analysis Plan

Document Number:

Document Version: 3.0

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:min) - UTC timezone
[redacted]	Clinical Approval	17-Jul-2020 13:48:50
[redacted]	Biostatistics Approval	17-Jul-2020 03:09:42
[redacted]	Clinical Pharmacology Approval	20-Jul-2020 15:06:29